INTEGRATED DATA ANALYTICS OF GERMLINE MUTATION CLASSES IN HUMAN CANCERS

M. H. AL-SHAMMARI

PhD

2013

INTEGRATED DATA ANALYTICS OF GERMLINE MUTATION CLASSES IN HUMAN CANCERS

An Integrated Bioinformatics Analysis to Investigate Associations between Germline Mutation Classes and Human Cancers

Mohamad Hilal Al-Shammari

BEng (Hons)

Submitted for the degree of
Doctor of Philosophy

School of Computing, Informatics and Media

University of Bradford

2013

Abstract

Integrated Data Analytics

of Germline Mutation Classes in Human Cancers

An Integrated Bioinformatics Analysis to Investigate Associations between

Germline Mutation Classes and Human Cancers

Keywords:

Cancer, Mutation, Gene, Chromosomes, Pathways.

Biological and environmental factors contribute collectively to the development of human cancers. The primary focus of this research project was to investigate the impact of germline gene mutations, as a significant biological factor, on 29 major primary human cancers. For this I obtained data from multiple databases, including the Genetic Association Database (GAD), Sanger database (COSMIC), HGMD database, OMIM data and PubMed literature. Using the Extraction Transform and Load (ETL) process, 424 genes were obtained with 8,879 cancer mutation records. By integrating these gene mutation records a Human Cancer Map (HCM) was constructed, from which several sub-maps were derived based on particular mutation classes. Furthermore, a Protein-Protein Interaction Map (PPIM) was constructed based on the encoded proteins of the 424 gene set.

Several key questions were addressed using the HCM and its sub-maps including the following: (i) Are individual groups of primary cancers associated with specific subset of genes (within the 424 full set)? (ii) Are groups of primary cancers associated with particular mutation classes? (iii) If both questions prove to be true, are groups of cancers associated with particular mutation class of target genes? This project also explored whether a corresponding Protein-Protein Interaction Map, derived from the Missense/Non-sense Mutation portion of the HCM gene set, would provide further information on gene associations between primary cancers in terms of the consequent identical amino acid changes involved.

Results showed that: (1) closely-connected human cancers in the HCM exhibited a strong association with a particular mutation class; (2) Missense /Nonsense and Regulatory mutations played a central role in connecting cancers (i.e. via primary nodes) and so significantly influenced the construction of the HCM; (3) Genes with Missense/Nonsense and Regulatory mutations tended to be involved in cancer-associated pathways; (4) Using the kappa test to measure the extent of agreement between two connected primary cancers in the sub-HCMs, BRCA1, BRCA2, PALB2, MSH2, MSH6, MLH1, CDKN2A, and TP53 showed

highest agreement for 5 of 10 mutation classes; (5) From the PIPM, it was evident that BRCA1, MSH6, BARD1, TP53, MSH2 and CHEK2 proteins best connected Breast, Ovarian, Prostate and Bowel primary cancers, and so the latter could represent ‘driver proteins’ for these cancers.

In summary, this project has approached the analysis of gene involvement in human primary cancers from the starting position of the mutation class that harbours the specific gene mutation. Together with their downstream resultant alterations in the associated proteins, this analysis can provide insights into the relatedness of primary human cancers and their potential gene hierarchies. These data may therefore help us to understand more fully the etiology, diagnosis and potentially personalized treatments for cancer.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ALA	Alanine
ARG	Arginine
ASN	Asparagine
ASP	Aspartate
BL	Bladder Cancer
BOW	Bowel Cancer
BRA	Brain Cancer
BRE	Breast Cancer
CE	Cervical Cancer
Chr	Chromosome
CYS	Cysteine
GLN	Glutamine
GLU	Glutamate
GLY	Glycine
HCM	Human Cancer Map
HIS	Histidine
HN	Head and Neck
ILE	Isoleucine
KI	Kidney Cancer
LE	Leukaemia Cancer
LEU	Leucine
LU	Lung Cancer
LY	Lymphoma
LYS	Lysine
ME	Melanoma
MET	Methionine
MY	Myxoma
OC	Ocular Cancer
OE	Esophageal Cancer
OV	Ovarian Cancer
PAN	Pancreatic Cancer

PHE	Phenylalanine
PR	Prostate Cancer
PRO	Proline
SER	Serine
CM-HCM	Classes Mutations of Human Cancer Maps
SN	non–melanoma Skin Cancer
ST	Stomach Cancer
STS	Soft Tissue Sarcoma
TER	Termination
TH	Thyroid Cancer
THR	Threonine
TRP	Tryptophan
TYR	Tyrosine
VAL	Valine
WI	Wilma’s Tumour
WO	Womb Cancer

Acknowledgements

Firstly, I would like to mention that I'm deeply grateful of having Dr Yonghong Peng and Professor Desmond J Tobin as my supervisors.

Dr Peng has always encouraged and supported me throughout the hard work during my PhD study. He has been always inspiring me in research and very generous with his time, ideas and knowledge.

Professor Tobin has been providing his great expertise in biological science to drive this interdisciplinary project forward. Professor Tobin has always provided a valuable advice and very constructive discussions on the project. He also helped me attend the System Biology Europe conference 2012; without his contribution I would not been able to attend this conference. With the valuable inputs from my supervisors, I have been awarded the best poster in System Biology Europe Conference 2012, of which I am very grateful.

Massive thanks to my great mother, sisters and brother, for their strong support during the PhD course.

I would like to thank, from the bottom of my heart, my wonderful wife Mrs Sarah Al-Taie for her support and belief in my ability to do this research and for stopping me from worrying so much when things didn't go plan. Particularly I would like to thank my uncle Mr Zahaa Redha who financially supported me through my study.

I also would like to thank many of my friends who has been supported me, particularly to Mr Karim Hammoud and Ms Rona J Wilson who motivate me to do a PhD.

Chapter One

1. Fundamentals of Genes in Health and Disease

Until recently the approach most widely used to investigate the fundamentals of human genetic variations in relation to the progression of human health and diseases has been focused primarily on identifying the variations of an individual candidate gene with the described patterns of Mendelian inheritance disease. The origin of Mendelian patterns of inheritance disease was first proposed by Gregor Mendel in 1865, on the basis of dominant and recessive genes (Mendel, 2008). Humans carry two copies of the same gene, each one of these copies is inherited from both parents. If one of the inherited gene copies is mutated on one of the autosomal chromosomes (chromosomes 1-22) it can lead to the development of diseases, such as Huntington's disease (Griffiths, 2005), neurofibromatosis type 1,2 and hereditary non-polyposis colorectal cancer etc. This is known as autosomal dominant. if both copies of the inherited genes from both parents are mutated then this can be called autosomal recessive, which is evident in newly born babies and certain rare conditions like cystic fibrosis (Flotte and Laube, 2001), sickle-cell disease and Tay-Sachs disease. Otherwise, if one copy of the mutated gene

occurs on the X chromosome, this can be called ‘X-linked dominant’ when the mutated gene is acting in dominant manner. The impact of the mutated gene can have influence on causing a disease in both males and females (as both genders have an X-chromosome), like Incontinentia pigment disease (Meuwissen and Mancini, 2012). If both gene copies are mutated this is known to be ‘X-linked recessive’ leading to disorders including Duchenne muscular dystrophy (Moat et al., 2013), or the less serious red-green colour blindness (Lenaers et al., 2012) and male pattern baldness. Finally, if the mutated gene occurs on the Y chromosome, this can be called ‘Y-linked’, which is implicated for males only, and is known to be associated with male infertility (Yang and Zhang, 2010).

Secondly, genetic heterogeneity influenced by environmental factors, i.e., etiologic heterogeneity, is associated with multiple groups of diseases that occur in populations e.g., breast cancers (Ford et al., 1998). Identifying the etiologic heterogeneity can be an important step toward detecting the possible impacts of gene-gene and gene-environment interactions on disease progression. A third approach for consideration is to identify the influence caused by a group of genes on each other leading to the presentation or expression of diseases. Systems biologists have shown that diseases that share phenotypic similarity are often associated with genes with strong functional associations (Oti et al., 2008). Also, genes associated with similar functions tend to have similar expression patterns (Chu et al., 1998), which provide us with an opportunity to gain insight into the functional association of genes. These valuable insights have opened up the potential for a virtual roadmap for physicians, genetic

counsellors and biomedical researchers for the further study of disease associations and will help our understanding of how diseases are caused and so may be successfully treated.

1.1. Chromosomes

There are 23 pairs of different chromosomes in human cells, with one member of each pair inherited from each parent. One of these pairs represents the so-called sex chromosomes, with either two X or one X and one Y chromosome. Only the father can contribute the Y chromosome to produce a male offspring, while both parents can contribute an X chromosome to produce a female child discussed in (Pasternak, 2005).

Eukaryotic chromosomes are mixtures of DNA and proteins, which together are called chromatin (Figure 1.1). A small amount of RNA is also present. Two classes of proteins, called histones and non-histone (or acidic) proteins are present to play a central role in chromatin structure. Proteins that form histones have a highly conserved amino acid sequence, suggesting that these molecules are very important. The basic building block of chromatin structure is called the nucleosome (Parry, 2001; Brown, 1992).

Nucleosomes consist of a core of histones around which DNA is wound. The core of DNA consists of two discs arranged in parallel, each composed of four histone molecules. The linker length between the nucleosomes varies. In humans it is about 60 base pairs (bp), giving a

total length of DNA per nucleosome of approximately 200 bp (Parry, 2001; Brown, 1992).

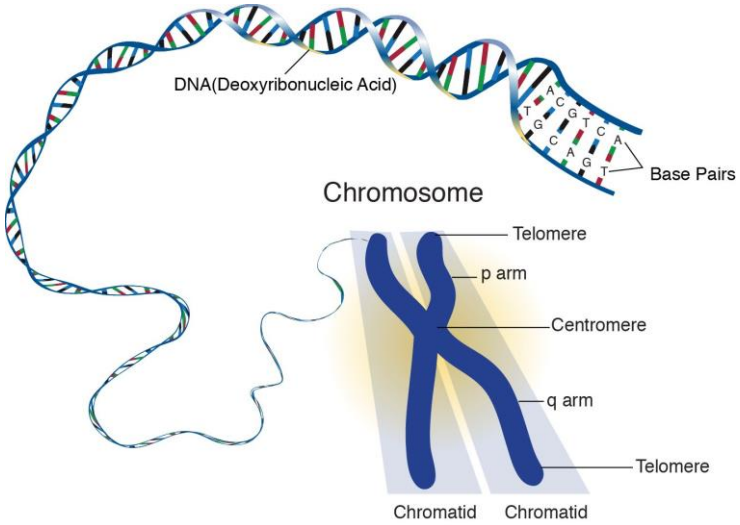


Figure 1.1: An Illustration of a Chromosome. Each chromosome is divided into two parts by a point known as a 'Centromere'. Chromosomes consist of four arms, two is known as short arms, labelled as 'p,' and the other two known as long arms, labelled as 'q'. From (Leja, 2010)

1.2. DNA

The DNA molecule consists of three basic components: pentose sugar, (i.e, deoxyribose) and a phosphate group, as well as four types of nitrogenous bases, of which cytosine and thymine are single carbon-nitrogen rings (called pyrimidines), and adenine and guanine are double carbon-nitrogen rings (called purines). It is important to note that the four bases combine with hydrogen ions in acidic solutions (Parry, 2001; Roberts and Pembrey, 1970; Brown, 1992). The chemical structure of the four bases that constitute the DNA molecule is shown in figure 1.2. There

are three hydrogen bonds between the cytosine-guanine pairs, whereas there are two hydrogen bonds between the adenine-thymine pairs (Parry, 2001; Roberts and Pembrey, 1970; Brown, 1992).

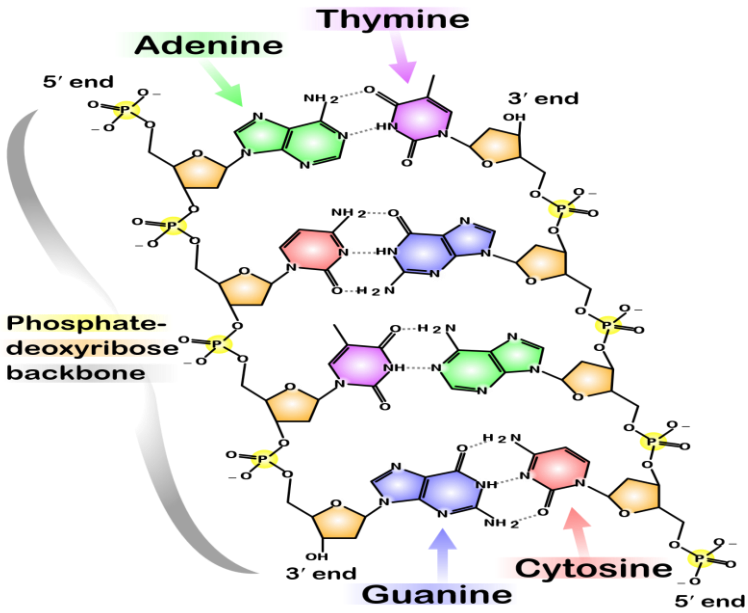


Figure 1.2: Chemical structure of the DNA molecule. From (Wikipedia, http://en.wikipedia.org/wiki/File:DNA_chemical_structure.svg).

The double helix DNA molecule contains sugar-phosphates, which are held together in order to form the backbone of the polynucleotide strands. It is important to note that the polynucleotide chains run in opposite directions so that the polynucleotide has a free 5' phosphate at one end and a free 3' OH at the other end. The sequence of bases encoding the genetic information is read as 5' to 3' or 3' to 5'. The four nitrogenous bases face the inner surface of each strand; hydrogen ties the four nitrogenous bases in the DNA strands, which creates the double helix. Each DNA subunit consists of one deoxyribose, one phosphate group and one base, called a nucleotide. Different sequences of nucleotide bases (ACCAAGTGC) specify a code for different proteins. Every 10 nitrogenous bases is executed to form the double helix. The sequence of base paring is

restricted such that adenine can only pair with thymine and guanine, can only pair with cytosine. This is called *Complementary Base Pairing*. This allows genetic information to be preserved during replication of the DNA and expression of the gene (Parry, 2001; Roberts and Pembrey, 1970; Brown, 1992; Lodish et al., 2000), as shown in figure 1.3.

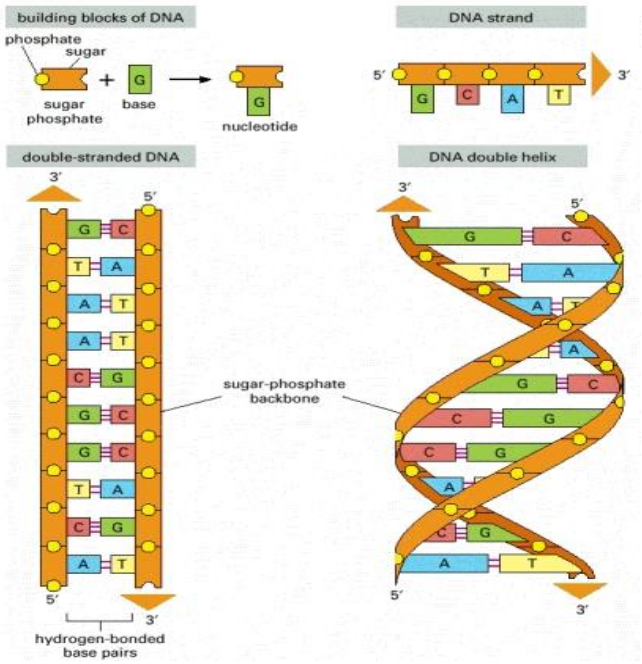


Figure 1.3: DNA and its building blocks. From (Lodish *et al*., 2000).

1.3. Genes

The DNA provides biological information for the organism, which will be encoded in a base sequence of the DNA and structured as an enormous number of genes. Individually these contain the instructions for the synthesis of a polypeptide (i.e. protein). From a physical perspective, a gene is a distinct segment of DNA with a base sequence that instructs the amino acid sequence of a polypeptide (Brown, 1992).

Genes differ significantly in size (the number of base pairs), ranging from less than 100 base pairs to several million base pairs. In higher-level organisms, the genes are run in a sequence of extremely long DNA molecules, named chromosomes. In humans there are around 20,000 to 25,000 genes, organized in 23 chromosomes (Quackenbush, 2011).

Biological information is carried by one of the two strands of the DNA double helix, which is known as the template/non coding/antisense strand, and can be used to generate an RNA molecule of the matching sequence, to direct the synthesis of the corresponding polypeptide. Whereas the term 'gene' is often applied to the non-template/coding/sense strand, which allow the RNA polymerase reads and transcribes into mRNA, for a transcription into mRNA to occur a promoter sequence must be in the correct orientation position (Pasternak, 2005).

1.4. Genetic Code

The genetic code, as shown in Figure 1.4, illustrates how base sequences are transformed into amino acid sequences, where the DNA sequence of a gene is separated into units of three nucleotides. A nucleotide is composed of a base, a five-carbon sugar (either ribose or 2'-deoxyribose) and a phosphate group. Each triplet nucleotide is called a codon and is indicative of a particular amino acid. Exceptions include methionine and tryptophan, which can be encoded by more than one codon (Figure 1.4).

Codons that specify the same amino acid are called synonym codons and tend to be very similar; for example ACU, ACC, ACA and ACG, where the

changes are all at the third position of the codon - called the "wobble position" (Brown, 1992; Hames and Hooper, 2000). Uracil is one of the 4 bases in the nucleic acid of RNA.

The genetic code is said to exhibit degeneracy. This is a description of redundancy, but without ambiguity. For example, although codons GAA and GAG both specify glutamic acid (an example of redundancy), neither of them specifies any other amino acid (therefore no ambiguity). Degeneracy of the genetic code minimizes the effects of mutations so that alterations to the base sequence are less likely to change the amino acid encoded with possible deleterious effects on protein function. It is important to note that there are 64 possible codons. 61 of these encode amino acids, while the remaining three codons (TAA, TAG, and TGA) act as signals for protein synthesis, which are known as termination codons or stop codons. The codon for methionine signals protein synthesis to start, and is known as the initiation codon (Brown, 1992) (Figure 1.4).

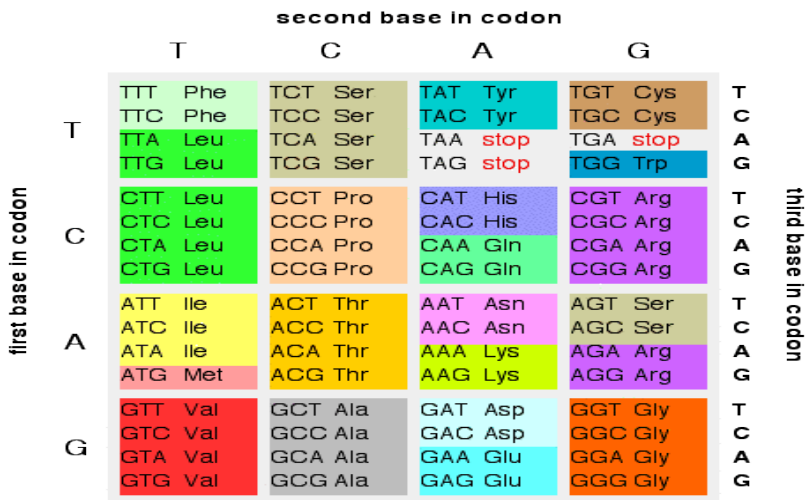


Figure 1.4: Genetic codes, where the letters T, C, A and G are the actual base letters and the abbreviation letters after the codons is the produced amino acid from each codon (Brown, 1992).

1.5. Mutations

The term **Genotype** describes the genetic or specific allele makeup of an individual and so also can inform the mutation types of specific genes under assessment. The **Phenotype** describes the individual's observable characteristics or traits (e.g. morphology or other biochemical or physiological properties, behaviour etc.). Thus, genetic mutation can result in phenotypic change. Different forms of mutations include: ***Regulatory, Complex rearrangements, Gross deletions, Gross Insertions, Missense/Nonsense, Repeat Variations, Small deletions, Small indels, Small insertions and Splicing variants (Joshi, 1997)***. These genetic mutations can give rise to several possibilities of human diseases when function is negatively affected.

Mutations can occur due to errors during DNA replication or be caused by environmental factors, such as a chemical mutagen, cosmic radiation, chemical carcinogenic, virogenic and ionic radiation. Environmental factors can have effects on human health and cause variance at the gene level, especially in cases of somatic cell mutation (Figure 1.5). This figure illustrates the intrinsic factors that cause genetic mutations in the human genome; either via an error occurring during DNA replication or methylation or via a transposable genetic element. By contrast, extrinsic factors (e.g., physical, chemical and cosmic radiation) can also cause gene mutations.

Besides errors in a DNA replication, gene mutations can be caused by various events: -

- If the gene-coding region is affected this could lead to alteration of the amino acid sequence in the resultant polypeptide chain of a specific protein (Kimball, 2003).
- Methylation, known as the biochemical process for adding the methyl group to the cytosine or adenine DNA nucleotides. Whereas the Transposable Genetic Element is the DNA sequence that can change its position within the genome. These two terms can also give rise to abnormal gene products, which can behave like gene mutations. However both sometimes only provide temporary prevention of normal gene expression, that leads to modification of phenotypic performance, known as the epigenetic effect (Kimball, 2003).
- Epigenetics is the modification to the genome of cancer cells, which is caused by a number of factors (e.g., methylation), and usually occurs due to methylation at the cytosine base (CpG). The epigenetic change does not involve nucleotide sequence changes, and it is as important as genetic mutations in a cell's transformation to cancers (Novak, 2004).

Furthermore, a gene mutation is a change in DNA sequence away from 'normal' to abnormal genetic variance. In contrast, a polymorphism is a common variation in DNA sequence among individuals in a population. To distinguish a mutation from a polymorphism, there is an arbitrary cut-off point observed as ≥ 1% prevalence in a population. If this found to be < 1% then the allele is regarded as a mutation (Ahmadian et al., 2000).

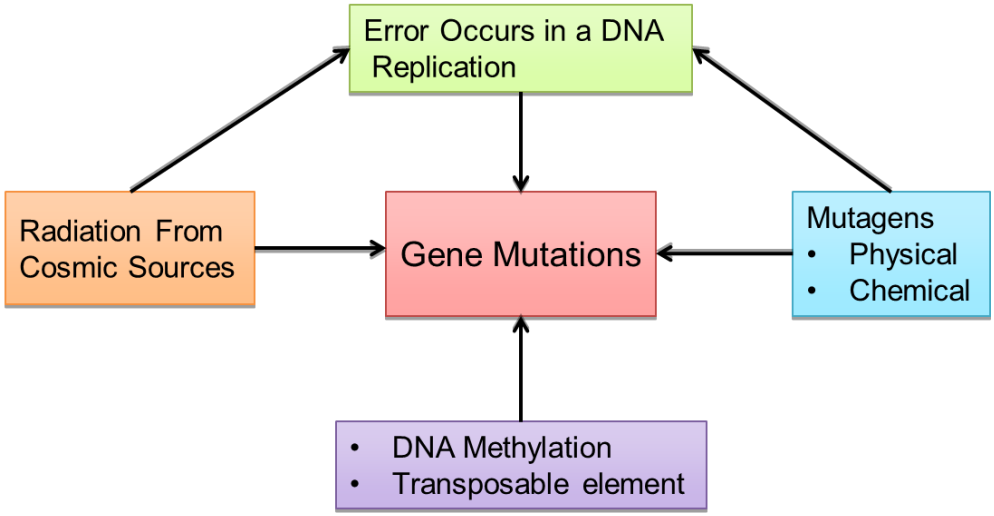


Figure 1.5: Causes of Genetic mutations.

1.6. Gene Mutation Classes

1.6.1. Frame Shift Mutation (Insertions and Deletions)

A frame shift mutation involves the deletion or insertion of a single nucleotide, which results in extra or missing amino acids in a polypeptide chain of a protein, with the possibility of having a deleterious effect. Deletions and insertions tend to be especially harmful when the number of missing or extra base pairs is not a multiple of three. This is due to the fact that codons consist of groups of three base pairs and such insertions or deletions can alter all of the downstream codons (Bell et al., 2003). See the example figures 1.6 and 1.7 for single base insertion and deletion.

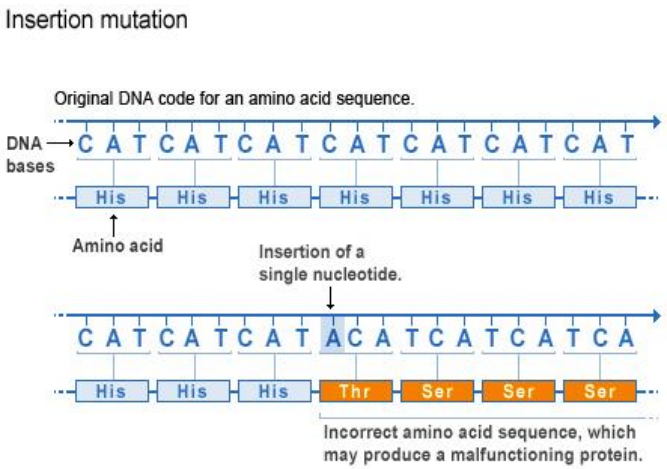


Figure 1.6: The insertion of single base (an A in the second codon) would convert the DNA sequence so it reads as 5'-CAT CAT CATCATCATCAT CAT-3' To 5'-CAT CAT CAT ACA TCA TCA TCA-3'. This would change the amino acid sequence from His-His-His-His-His-His-His to His-His-His-Thr-Ser-Ser-Ser. *Image from (Medicine, 2013b)*

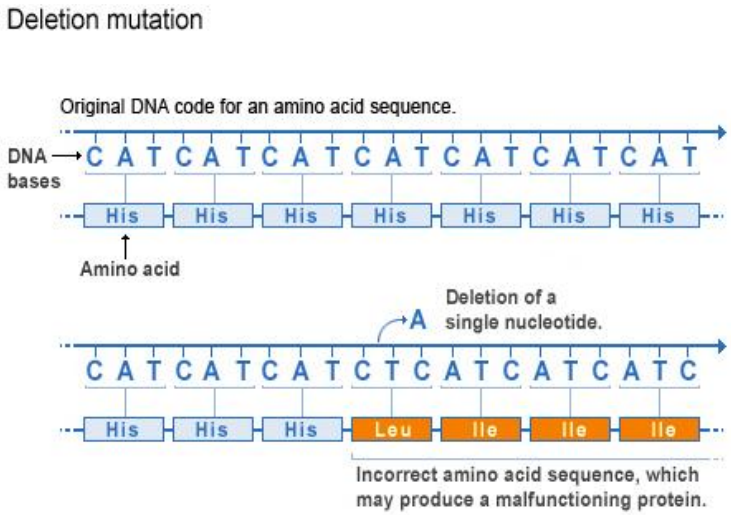


Figure 1.7: The deletion of single base (an A in the second codon) would convert the DNA sequence so it reads as 5'-CAT CAT CATCATCATCAT CAT-3' To 5'-CAT CAT CAT CTC ATC ATC ATC-3'. This would change the amino acid sequence from His-His-His-His-His-His-His to His-His-His-leu-Ile-Ile-Ile. *Image taken from (Medicine, 2013a)*

1.6.2. Missense/Nonsense Mutation

Missense mutations involve a change in a single amino acid that may affect a protein's structure or function, and is thus likely to lead to a harmful effect in the individual and to generate a mutant phenotype (Kozak, 1986). Nonsense mutations can produce, in error, one of three stop codons (UAA, UAG or UGA) in messenger RNA (mRNA). The stop codons can terminate both mRNA transcription and translation and so result in premature termination of the polypeptide chain. On the other hand, if a correct stop codon is altered to now encode for an amino acid, an abnormally elongated polypeptide is then produced (Welch et al., 2007) (Figure 1.8).

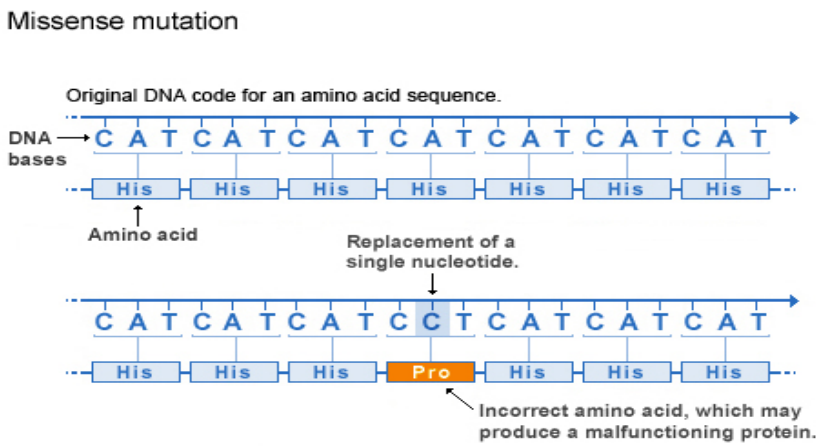


Figure 1.8: Missense mutations, of single base (an A in the 4th codon) would convert a DNA sequence so it reads as 5'-CAT CAT CAT CAT CAT CAT CAT-3' to 5'-CAT CAT CAT CCT CAT CAT CAT-3'. This would change the amino acid sequence from His-His-His-**His**-His-His-His to His-His-His-**Pro**-His-His-His. *Image Taken from (Medicine, 2013c)*

1.6.3. Splice Site Mutation

Various types of RNA molecules play key roles in making proteins. The gene transcript (mRNA) transfers information from **DNA** in the nucleus to the **ribosome** in the cytoplasm that makes protein. The mRNA molecule is used in protein formation via a process called RNA splicing via the excision of Introns (non-coding elements) and re-connecting the Exons together (see figure 1.9). Arranging Exons in different patterns, called alternative splicing, enables the cells to make different proteins from a single gene (Norvell and Machalek, 2000). Splicing has to be extremely accurate; any errors in the splicing process, even one that results in the deletion of just one nucleotide in an Exon or the addition of just one nucleotide in an Intron, will throw the whole sequence out of alignment. The result is usually an abnormal protein or no protein at all (Norvell and Machalek, 2000). For example, splicing error mutations are the causative factor of one form of Alzheimer's disease (Kinzler and Vogelstein, 1998).

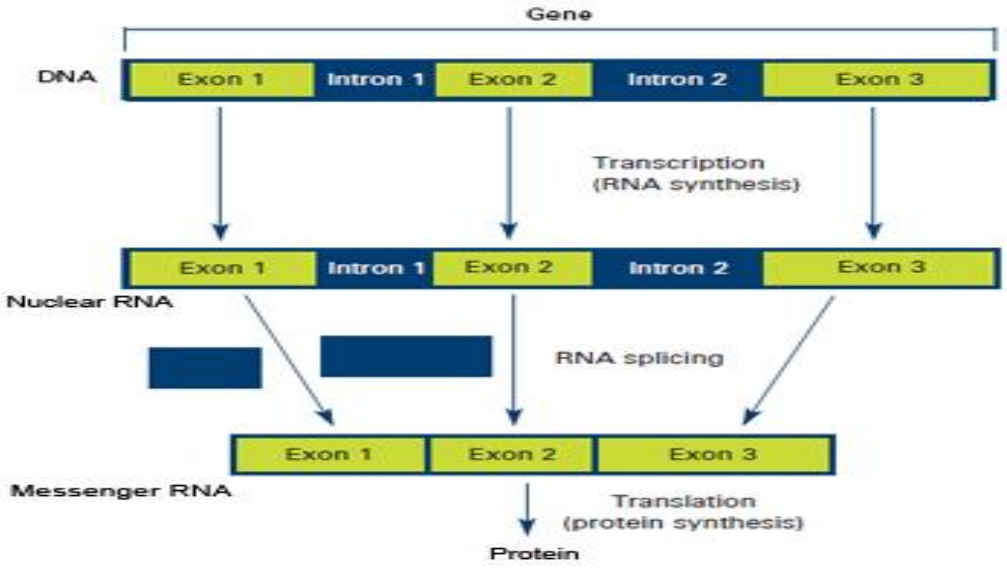


Figure 1.9: Mutation of normal Splicing *(Norvell and Machalek, 2000)*.

1.6.4. Repeat Variations Mutation

Genetic variation is important source of genetic material for natural selection. The natural selection is the result of the interactions between genetic variations in a population and the environment. A variation mutation in the alleles of genes can occur within and among populations, which can lead to a permanent change in the primary structure of a gene. However there are three primary factors of genetic variation (i) a single nucleotide change in the DNA sequence of genes, (ii) A transfer of a gene or alleles from one population to another i.e gene flow (gene migration). For example, the mixed populations of humans in the USA has proved the transfer of malaria resistant gene from the migrant population to the native population among whom the gene was absent originally , (iii) introducing a new gene combination into a population i.e sex and genetic shuffling. (Conrad et al., 2011)

1.6.5. Small Indels Mutation

The databases used in this study additionally defines a mutation class as ‘small indels’ and represent different records from those included in the frame shift mutation (see above). ‘Indel’ stands for mutations with either an insertion or a deletion, i.e., a mutation class that includes both insertions and deletions (Kondrashov and Rogozin, 2004; Ogurtsov et al., 2004). Indels differ from point mutations; where an Indel inserts and deletes nucleotides from a sequence, a point mutation is a form of

substitution that replaces one of the nucleotides (Figure 1.10). Indels can also be contrasted with Tandem Base Mutations (TBM), which may result from fundamentally different mechanisms (Hill et al., 2003). (See figure 1.10).

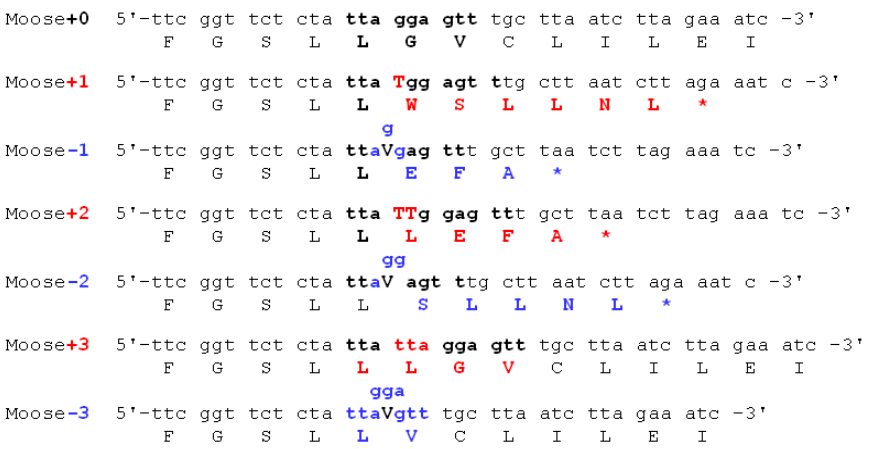


Figure 1.10: Small Indels, where line 1 shows the standard moose sequence, with 5th, 6th, & 7th triplets highlighted, Lines 2 & 3 show a **single-nucleotide insertion** (**T**) or **deletion** (**g**) at the sixth triplet, Lines 4 & 5 show a **double-nucleotide insertion** (**TT**) or **deletion** (**gg**) at the sixth triplet and Lines 6 & 7 show a **triplet insertion** (**TTA**) or **deletion** (**gga**) of the sixth triplet (Hill et al., 2003).

1.6.6. Regulatory Mutation

Cis-regulatory element is a region of DNA or RNA that regulates the expression of genes located on that same DNA molecule. The word 'cis' means "on the same side as" (Wray, 2007). Regulatory mutations represent a significant cause of human disease. According to April 2009 statistics compiled by the Human Gene Mutation Database, 1459 regulatory mutations have been identified in over 700 genes that cause human-inherited disorders. The majority of these regulatory mutations are located in proximal and distal promoter elements (a region of DNA that facilitates the transcription of a particular gene) that map within 1 kb of the transcript start site (TSS). Cis-regulatory mutations affect a broad range of morphological, physiological and neurological phenotypes (Epstein, 2009). (Figure 1.11).

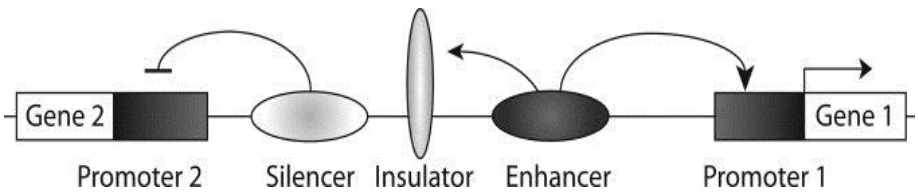


Figure 1.11: Regulatory Mutation: two different promoters are located next to the two different genes. By binding to Promoter 1, an enhancer actively regulates Gene 1 and leads to its transcription, as indicated by the black arrow above Gene 1. An insulator prevents this enhancer from activating Gene 2, which is not transcribed due to regulation by the silencer (Worsley-Hunt et al., 2011; Epstein, 2009).

1.6.7. Chromosome Rearrangement Mutations

1.6.7.1. Translocation

Chromosomal translocation, a chromosome abnormality or aberration caused by rearrangement of parts. The mutation is caused by a rearrangement of two different non-homologous chromosomes breaking and re-joining, where genes in one part of the one chromosome is replaced with another genes in part of different chromosome. A fusion gene might be creating as a result of the translocation which knows to be a common type for causing cancers (Kinzler and Vogelstein, 1998).

1.6.7.2. Complex Rearrangements (Inversion)

This class represents a process where a part of the chromosome is reversed end-to-end, and inversions are most likely to happen when individual chromosome suffer a break and are then rearranged. There are two types of Inversion, the first one can happen in one arm of the chromosome which knows to be as ‘Paracentric’ and the other one can happened in both of the chromosome arms with the involvement of the centromere breaking point which is knowing as ‘Pericentric’ (Kinzler and Vogelstein, 1998).

1.6.7.3. Genetic Duplications/Amplification

Gene duplication is a mechanism to generate new genetic material i.e., molecular evolution. Duplications occurs in chromosomes contains a proto-oncogene within them where the number of copies of proto-oncogene is increased in that chromosome. However these newly introduced copies of the proto-oncogene produce protein that is involved in stimulating cell growth. Therefore the amount of the produced proteins can be increased, which might lead to the cause of cancers (Kinzler and Vogelstein, 1998).

1.6.7.4. Gene Deletion/Deficiencies

Genetic aberration is the process where a part of chromosome or DNA sequence is deleted resulting a loss in a genetic material. The deletion could be errors in the exchange of genetic material between homologous chromosomes (crossing over) during the cell division (meiosis), which can cause serious disorders such as cancers (Kinzler and Vogelstein, 1998).

1.7. Literature Review of Gene Mutations in Human Cancers

Various attempts and approaches have been made in order to understand genetic mutations classes and the related human cancers arising from these mutations. A review of the most cited literature studies related to this field is discussed in this section.

1.7.1. Relationship between Gene Mutations Classes and Human Cancer

During the past decade biomedical researchers have examined a small number of cancers sharing commonalities in both etiology and pathology. The latest growth in genetics and genomics studies has enabled us to interrogate the entire genome for the influence of genes in causes of human cancers (Jimenez-Sanchez et al., 2001; Cooper et al., 2010). These impressive studies have opened the gates wide for cancer genomic research to investigate the implications of mutation classes in human cancers.

Many authors have been able to identify cancers based on mutation classes, starting with the Missense/Nonsense mutations, where detection of a single nucleotide changes leading the genetic sequence to encode for different amino acid. For example in studies of (Sweet et al., 2010; Sluiter and van Rensburg, 2011; Bonadona et al., 2011) some cancers (such as Breast, Lung and Bowel) were considered as a consequence of altered amino acid, that occurred due to dozens of single base-pair substitutions,

in a group of genes (e.g., *ATM, BRCA1, BRCA2, MSH2, MSH6* and *MLH1*).

Others have detected how a change of nucleotides during splicing of an intron could transform the precursor messenger RNA into mature messenger RNA, leading to unusual production of proteins. For example, altered distributions of the exonic splicing enhancer motif in ATM gene resulted in a change in protein function that are associated to breast cancer (Brunet et al., 2008; La Paglia et al., 2010).

Remaining with the single base-pair substitutions, a regulatory mutation in the FAS gene promoter affects gene expression and modulates apoptotic signalling and can cause Acute Myeloid Leukaemia (Sibley et al., 2003). Yang and co-workers examined 1,840 lung cancer patients and found two single nucleotide polymorphisms (SNPs) located in the FEN1 promoter, which causes the reduction of *FEN1* expression level, leading to the development of lung cancer (Yang et al., 2009).

Small insertions and deletions can alter human genetic sequences and can result in cancer formation. One of the most common human genetic cancers, breast cancer, is frequently caused by a small coding insertion or deletion within the *MSH2, MSH6* and *MLH1* genes that eliminates a single amino acid (Kim et al., 2010; Nilbert et al., 2009). Another form of deletion and insertion, include gross deletions for the entire *FH* gene and gross insertions or duplications for the *FH* gene in two different patients; both processes can cause kidney cancer (Bardella et al., 2011; Ahvenainen et al., 2008). Others (Vaughn et al., 2010) have further proposed how the whole deletion of the Mismatch repair gene (PMS2) in 22 out 53 patients is

principally responsible for a form of bowel cancer (Walsh et al., 2010). Discovery of the inheritance patterns for mutations of large rearrangements (Gross Insertion) in the *BRCA1* gene has revealed the importance of this gene in breast cancer. A study of large genomic rearrangements (complex rearrangement mutations) in breast cancer cases (Engert et al., 2008), has revealed that *BRCA1* can harbour cancer-causing mutations. An analysis of 1,506 German families carrying breast and ovarian cancers detected a total of 32 pathogenic rearrangements in the *BRCA1* genes. *BRAC1* gene screening has been carried out on 283 ovarian cancer patients to detect large genomic alterations, with the finding that 12 of these patients exhibit complex rearrangements mutations (Ramus et al., 2007).

A small indel have also been found to be a cancer-causing mutation. It tends to have a deletion of a single base or double nucleotide insertions affecting the genetic sequence, such as the finding of a small indel mutation in *HLA-DQA1* causing lung cancers (Qiu et al., 1996). Finally, a reported study of the *MUC6* gene, which is a highly expressed gene in stomach and gall bladder, has been carried out on 470 gastric cancer patients, which showed that five of these cancer patients are known to have repeat variations mutations in this gene (Kwon et al., 2010).

1.7.2. The Analysis of Genetic Data in Cancer Research

The past few decades has witnessed a rapid development in technologies that can analyze cancer genomes. These have focused on a single platform or type of genetic alteration such as High Resolution Melt (HRM) analysis for the detection of novel oncogenes (Nanjundan et al., 2007). Similarly, the technology of DNA Sequencing (*i.e.,* directed sequencing, shotgun sequencing) has been used widely to detect novel genes involved in specific cancer types e.g., direct sequencing to detect mutations in *BRAF* causing human cancers (Davies et al., 2002). More recently, a combination of genome sequencing and bioinformatics tools has been used, e.g., by The Cancer Genome Atlas (TCGA), to catalogue genetic mutations associated with cancers. All these developments have opened the road widely for new genetic analysis tools. For example, a new technique has been implemented to sequence DNA, named as the Second-generation sequencing. This new technique offers several advantages over previous technologies; (i) allows the whole genome to be sequenced, (ii) offers structural information, which has never been available before using other platforms, (iii) enables for the first time a global assessment of chromosomal rearrangements in cancer. A good example of next generation sequencing can be found in (Mardis et al., 2009).

1.8. Method for the Analysis of Biological Networks

A network is a graphical representation of relations between objects, where these objects potentially interact with each other. In human genome research, graphs or networks have been used to represent various complex biological processes such as regulatory networks, disease networks, and various types of pathway interactions such as protein- or gene-interactions. The development of molecular networks was begun over 40 years ago using classical molecular tools (Dagley and Nicholson, 1970). With the rapid increase genome data, scientists now are able to construct more informative molecular networks, such as protein-protein interaction networks and gene regulatory networks. These molecular networks have similarities in terms of their structures. For example, the yeast transcription regulatory network (Guelzim et al., 2002), the metabolic networks (Wagner and Fell, 2001), and the yeast protein interaction network (Wagner, 2001), all have approximately power-law degree distribution, where most nodes in the networks interact with one other node, and a few can interact with tens or hundreds of others. Various mathematical approaches have been used to investigate the structure of biological networks including (i) degree distributions of nodes (k), (ii) clustering coefficient. The degree distribution (Newman, 2003) is used to quantify the properties of each node connecting to other nodes within the network. The clustering coefficient is to observe the connections of a node and connections made by its neighbours, this observation can

only be applied to a large fragment of the network (Newman, 2003). See figure 1.12

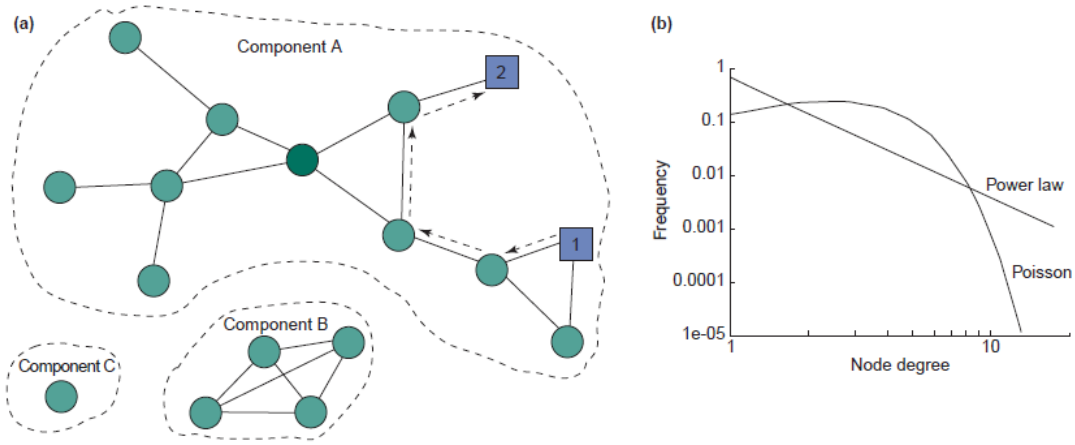


Figure 1.12: Theoretic illustration of a network (Proulx et al., 2005) (a) divided into three components A, B and C outlined with dash line. Each component node is represented by circles or squares. These nodes are connected via edges (solid black lines). The degree of connectivity for each node varies between 0 and 4, for example the central node with degree 4 is shaded darker green. In component (A) the shortest path between square node 1 and square node 2 consists of 4 steps. Whereas component (B) has a high clustering coefficient, due to the connectivities of its neighbour (i.e, every set of three nodes is connected within a triangle of edges). (b) Comparison of the node degree and frequency distributions produced by models of network formation, displayed on a log-log plots. Poisson random networks and Power-law degree distributions that appear linear are shown in the Figure 1.12 (Proulx et al., 2005).

1.9. Aims and Objectives

The main aim of this study is to investigate the influence of various germline mutation classes on different human cancers, by means of (i) constructing a Human Cancer Map (HCM) based on a knowledge of genes and on their associated mutation classes, (ii) investigating the distribution of mutation classes over the newly constructed Human Cancer Map (HCM), (iii) constructing a Genome-wide Distribution Map to identify the involved genes in the constructed Human Cancers Map, (iv) identifying pathways associated with the involved genes in the constructed Genome-wide Distribution Map, (v) identifying the relatedness of the involved genes in the Genome-wide Distribution Map based on their biological roles and, (vii) associating genes with cancer based on the corresponding encoded proteins and the protein-protein interactions (PPI) followed by their interrogation for Missense/Nonsense mutations.

This study was conducted in several phases: First, I investigated the properties of gene mutation classes in relation to human cancers; second, I gathered information about the ten classes of germline mutations involved. Third, I prepared the gathered biological data and established associations between gene mutations and human cancers. Thereafter, I build a human cancer map and a Genome-wide Distribution map based on the associations of specific mutation classes.

The objectives in order to achieve this aim included:

- Integrating the complementary data sources such as Genome database, OMIM, HUGO Gene Nomenclature Committee, Entrez

Gene, GeneCards, GenAtlas, GeneClinics, UniGene, SwissPort, Human Protein Reference Database, Human Gene Mutation Database (HGMD), Genetic Association Database and Sanger Database.

- Designing constructing and analysing networks using various bioinformatics tools including R and bioconductor, Cytoscape and Gephi network visualization tools. R is a programming language that has been widely used for statistical computation, graphics and bioinformatics. Cytoscape is an open-source bioinformatics software platform for visualizing molecular interaction networks and biological pathways and integrating these networks with annotations, gene expression profiles and other state data. Gephi visualization is a network analysis tool that has been used widely by bioinformatics.
- Designing and constructing a Genome-wide Distribution Map using CIRCOS visualizing software package (Krzywinski et al., 2009) and Perl Programming Language.

Chapter Two

2. Material and Methods

The aims of this chapter are four-fold: (i) to investigate the properties of genetics in relation to human cancers, by accessing various complementary data sources, including Genome Database (Grzybowska et al., 2002), Database of Human Genes and Genetic Disorders (OMIM) (Amberger et al., 2009), GeneCards (Safran et al., 2002a), Genetic Association Database (GAD) (Becker et al., 2004), Sanger Database (Futreal et al., 2004) and the Human Gene Mutation Database (HGMD) (Stenson et al., 2009); (ii) to extract genes associated with cancers and their associated classes of germline mutations including; *Regulatory, Complex rearrangements, Gross deletions, Gross Insertions, Missense/Nonsense, Repeat Variations, Small deletions, Small indels, Small insertions, and Splicing*; (iii) to extract the associated biological aspects of these data, with reference to Cancer Research UK and PubMed, (iv) to apply the appropriate statistical methods for analysing the data.

2.1. Data Integration

Data integration can be defined as a process of combining data obtained from different sources, with 'cleaning-up' the data so as to provide a unified view for users. This process is a significant in this study, as it involves complex data mining procedures to combine research results and findings from different depositories (Halevy, 2001). This process results in a new 'data warehouse', via procedures of Extracting data from various sources, and then Transforming it to construct a database suitable for a particular operation, and Loading it to be in a useable format (ETL) for data mining and analytics (Chaudhuri and Dayal, 1997). My aim here is to apply the ETL process to perform data extraction from three complementary data sources, including Genetic association database (GAD), Sanger database (COSMIC) and human gene mutation database (HGMD), and then transforming it (e.g cleaning, reformatting, standardization, aggregation) into a suitable data set as specific tables layout (See figure 2.1). The data described in the next two sections 2.2 and 2.3 will be manipulated based on the ETL process.

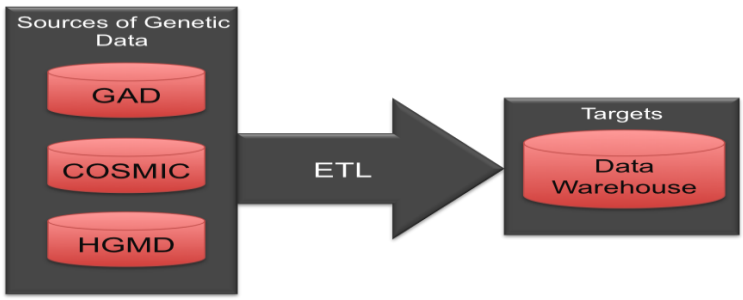


Figure 2.1: Extraction, Transform, and Load (ETL). On the left hand side is the list of three main biological databases to be used in this project. The ETL is the process to apply for the extracted data, and the data warehouse is when the data is ready to be suitable for the project needs.

2.2. Source of Data and Genetic Records Used in this Thesis

2.2.1. The Genetic Association Database (GAD)

The Genetic association Database (GAD) (Becker et al., 2004), is an online library archive of published genetic association studies that provides a comprehensive, public, web-based repository of molecular, clinical and study parameters for more than 130,000 entries of human genetic association studies. These includes 19 different disease classes, such as cancer, aging, cardiovascular, chemdependency, developmental, haematological, immune, infection, metabolic, mitochondrial, neurological, normal variation, pharmacogenomics, psychiatric, renal, reproduction, and vision. Each entry of the GAD is saved as an independent record and is composed of 22 fields (or attributes), including gene symbol, entrez GeneID, chromosomal location, associated tag between genes and disorders (“Y”, N), DNA position, P-value, reference and its corresponding PubMed is and OMIM id, etc.

The GAD data file was downloaded on 30 November 2011. Out of 21,444 cancer-related entries, 1,908 were selected based on their positive associations with cancers. The selected entries contained 486 unique genes and 480 duplicated primary or subtype cancer disorder names.

2.2.2. The Cancer Gene Census Database (COSMIC)

The Cancer Gene Census Database (COSMIC) (Futreal et al., 2004) is a collection of somatic and germline mutations with classes of mutations from the published literature. The mutation information is based on those mutations implicated in the development of particular cancers. The COSMIC database stores a total of 1,324 mutations entries. Table 2.1 summarise the characteristics of COSMIC data. Each entry of COSMIC is saved as a unique independent record and is composed of sixteen fields (attributes), including gene symbol, name of gene, geneID, chromosome, chromosome band, somatic mutation, germline mutation, tumour type (somatic mutation), tumour type (germline mutation), cancer syndrome, tissue type, cancer molecular genetics, mutation subclasses, translocation, other syndromes.

We downloaded the COSMIC database file on 1st of February 2012. Out of 487 cancer-related genes, 76 genes were selected with germline mutations.

Table 2.1: Total number of entries for each mutation type, mutation class, gene, and chromosome.

Category	Number
Amplification Mutation	16
Chromosome location	487
Frameshift mutation	100
Germline Mutation	76
Large deletions	37
Missense Mutation	141
Nonsense Mutation	92
Other mutation	26
Somatic Mutation	447
Splicing Mutation	63
Gene Symbol	487
Translocation Mutation	326

2.2.3. The Human Gene Mutation Database (HGMD)

The Human Gene Mutation Database (HGMD) (Stenson et al., 2009) is a large depository of broad data on human germline mutations with its sub-categories of mutation classes. The mutation data includes point mutations of a single base pair with insertions and deletions, regulatory and splicing-relevant regions of human nuclear genes, micro-deletions (indels), repeat variants, gross lesions (deletions, insertions and duplications) and complex rearrangements (including inversions, see table 2.2 for a summary of mutation data). The mutation data was stored as independent records and presented on a gene-wise basis, with access to the mutation classes data via a hypertext link, including additional data sources (i.e., Genome Database (GDB), Online Mendelian inheritance in Man (OMIM), HUGO Gene Nomenclature Committee (HGNC), Entrez Gene, GeneCards, GeneAtlas, GeneClinics, UniGene, SwissProt and the Human Protein Reference Database from each gene page.

Table 2.2: Total number of mutation classes held by HGMD [3]

Mutation type	Total of records
Single base-pair substitutions (Point mutation)	
Missense or nonsense	48,343
Splicing	8,219
Regulatory	1,400
Other lesions	
Small ($\leq$ **20** ***bp***) deletions	13,628
Small ($\leq$ **20** ***bp***) Insertions	5,567
Small ($\leq$ **20** ***bp***) Indels	1,244
Gross ($>$ **20** ***bp***) deletions	5,158
Gross ($>$ **20** ***bp***) insertions duplications	1,003
Complex rearrangements	736
Repeat Variations	260
Total	85,558

In order to extract data from the HGMD Web-based application a gene symbol is needed. The gene symbols used here were based on merging the GAD and COSMIC gene data, then performing a filtering process to remove duplicated genes. In total 520 unique genes were selected which represent the cancer-associated genes. The HGMD was then extracted manually by entering each of the gene symbols into the HGMD Web-based applications. The result of the search included gene symbol, gene descriptions, chromosome location, cDNA sequence ID, mutation classes, total number of mutations for each gene, total number of specific class of mutation for each of the gene, related phenotypes, unique mutation ID, PubMed references, and some additional information about the gene. I further expanded the search by using cancer names obtained from the GAD to increase the chance of finding new cancer-related gene and mutation records. As a result a total of 15,264 records were collected, with a combination of cancers and some non-cancers disorders for each of the 10 different mutation classes. Table 2.3 summarises the total number for each manually-curated mutation record.

Table 2.3: Classes of mutation; with total number of collected records for cancers and non-cancer disorders.

Class of mutation	Total number of mutation records
Missense nonsense	7456
Small Deletion	3280
Splicing	1465
Gross Deletions	1227
Small Insertions	1111
Regulatory	280
Small Indels	177
Gross Insertions	150
Complex Rearrangements	82
Repeat Variations	36
Total	15264

For further investigation the entries were cross-checked twice manually to ensure there was no duplication of the data and to double-check that the collected information is accurate and that none of the records were eliminated by the extraction process. As I was only interested in the cancer entries, a separation of the data on the basis of cancers and non-cancers entries was carried out in the data preparation part for each of the 10 classes.

2.2.4. DAVID Bioinformatics Database

Database, Annotation, Visualization and Integrated Discovery (DAVID) is an online functional annotation tools program. The tool aims to group a list of genes from different organism/species based on measuring the agreement of genetic biological functions (annotations). DAVID web-based tools tend to deliver other services, such as identifying the genetic pathways for a given list of genes, using bio-pathways, KEGG pathways and more (Sherman et al., 2007). My interest in DAVID online tools is to group our genes based on annotation associations using the functional classification tools.

2.2.5. The BioGRID Database

The Biological General Repository for Interaction Datasets (BioGRID) is a public database for genetic and proteomics interactions (http://thebiogrid.org/). The BioGRID contains 684,996 physical and genetic interaction records, of which 381,026 are about human cells (access date:01/10/2012). The BioGRID supplies an important plugin for Cytoscape visualization network tools, known as BioGRID Plugin, to allow

access of interaction data. This plugin was used in this research to search for the Physical Protein-Protein interactions related to the 424 selected genes (Stark et al., 2011).

2.3. Data Preparation

In the original HGMD, the 15,264-curated mutation records have much complexity with the data. For example the data was a combination of cancers and non-cancer disorder records. Also, many of the cancer entries are duplicated using different names/terms or the subtypes of cancer disorders are listed together with primary cancer names/categories etc. To deal with such data complexity, the non-cancer disorder records were eliminated from the list by investigating on a manual basis each of the disorders for the 15,264 records and the eliminations informed by the published literature studies (e.g., PubMed, OMIM and other online libraries) (Amberger et al., 2009). As a result 6,717 non-cancer records were eliminated from the 15,264 list (see table 2.4).

Table 2.4: The total number of eliminations for each of the mutation class for cancers and non-cancers disorders.

Mutation	Non-cancer Mutations	Cancers Mutations
Missense nonsense	4371	3085
Small Deletion	1181	2099
Small Insertions	374	816
Gross Deletion	325	902
Splicing	295	1091
Regulatory	91	189
Small Indels	34	143
Gross Insertions	19	131
Repeat Variations	16	25
Complex Rearrangements	11	66
Total	6717	8547

Secondly, the remaining 8,547 cancer records exhibited some naming/nomenclature complexity i.e., the primary cancer name had sub-classification names. For example, Leukaemia has two subclasses - Acute Leukaemia and Chronic Leukaemia, and each one of these a variety of subtypes. On the other hand many of the cancers were duplicated using 'different' names; e.g., Bowel cancer, Colorectal cancer and Colon cancer. These entries with duplicated names should be merged into the cancer term "Bowel cancer" etc. To solve this critical issue, a hierarchy list for each of the primary cancer names was manually created, followed by primary cancer subclass names and sub-subclass cancer names. In other words, constructing a family tree for each of the cancers, starting from primary, secondary, tertiary and so on. This family tree of cancers informed the re-arrangement of the data to be used for further development in the project in order to achieve our aims and objectives. See table 2.5 for an Illustration of the hierarchy of human cancers, constructed based on the terms applied in Cancer Research UK-accepted terms (Grzybowska et al., 2002). Moving further and to ensure no more complexity of the cancers disorder naming, I cross-checked each of the entries using the naming structure of table 2.5. This is to associate each cancer with its primary named cancer. While assigning each of the individual cancer entries into their 'primary' cancer designation (e.g, all cancers affecting the bowel designated as 'Bowel Cancer' as primary name), in some cases several cancers were found to be associated with the same mutation. For example, a Missense/Nonsense mutation in the *MLH1* gene (i.e. 'CGG' to 'TGG') resulting in an 'Arg' to 'Trp' change at

position 389 was found in both Breast and Colorectal cancer (Jensen et al., 2010), see table 2.6. Therefore, this mutation record was split into two, though both retained the same unique mutation ID. This approach allowed me to look at each of the cancer disorders as independent records when applying the statistical analysis for each gene mutation and cancer (see Tables 2.6 and 2.7). After the required data pre-processing (eliminations, merging and splitting) 8,879 mutation records were remained, containing 424 unique genes and a total of 26 primary unique cancer disorders. Each entry was stored as an independent record (see figure 2.2).

Finally, the data was manipulated to construct a table containing the following entries: gene symbol, unique mutation ID, sub class of mutation, cancers disorder name, total number of mutations for each gene, total number of mutations for each of the 10 different classes of mutations, and PubMed references (Table 2.8).

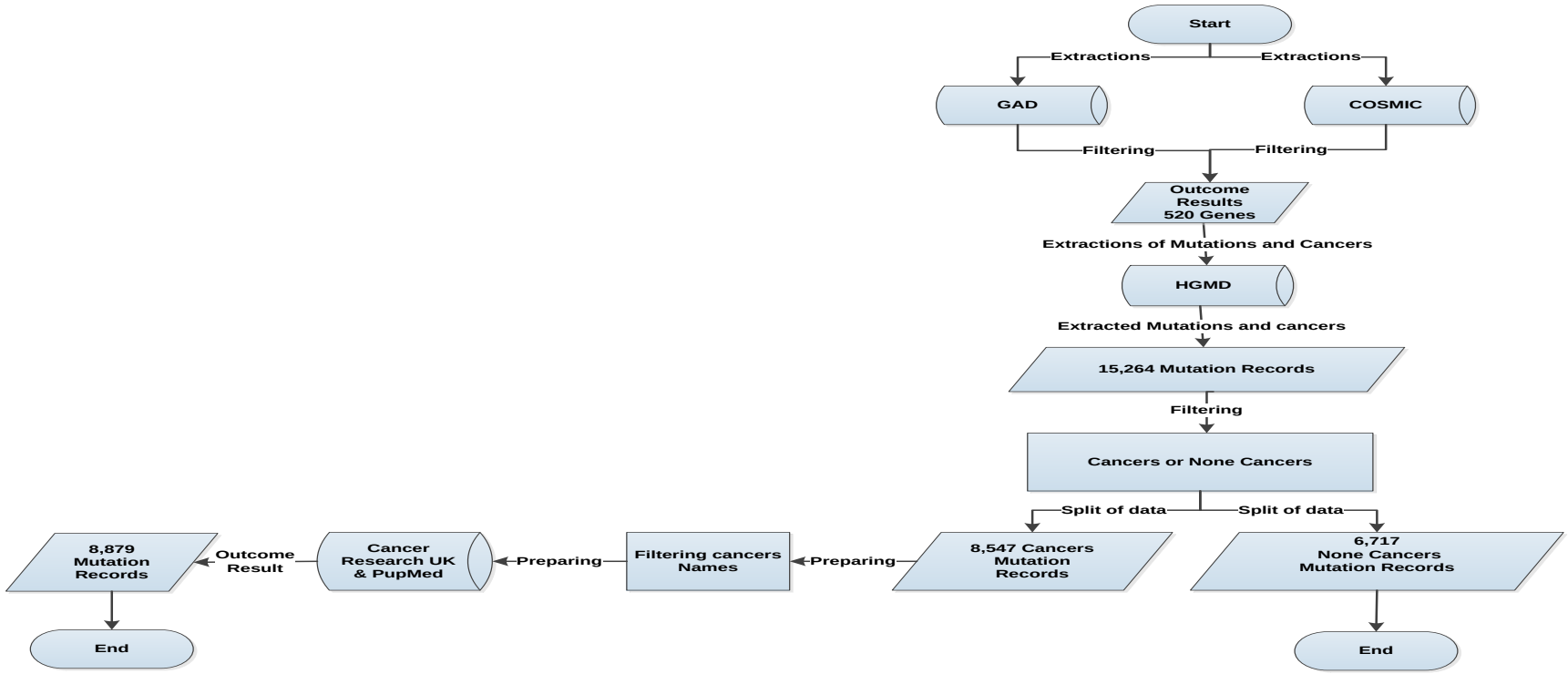


Figure 2.2: Algorithmic framework for the extractions and preparation of mutations data. The start point at the top of the image shows the two extracted databases for finding the cancer-associated genes. Moving to the middle of the image the HGMD database shows the extractions of mutation records, while moving down the image shows the split of the data for cancers and non-cancers. This is followed by filtering the data on the basis of primary cancer names.

Table 2.5: The hierarchy of human cancer disorders (family tree of cancers), where the red text represents the primary cancer name, the green text the subtype (secondary level) and the black text the tertiary level name. In total there are 29 primary cancer names, each one followed by its subtypes.

Hierarchy of cancers	
1	Leukaemia
1.1	Acute leukaemia
1.1.1	Acute Myeloid Leukaemia
1.1.2	Acute Lymphoblastic Leukaemia
1.2	Chronic leukaemia
1.2.1	Chronic myeloid leukaemia
1.2.2	Chronic lymphocytic leukaemia
1.2.3	Hairy cell leukaemia
2	Brain Tumour
2.1	Gliomas
2.1.1	Astrocytoma
2.1.2	Ependymoma
2.1.3	Oligodendroglioma
2.1.4	Mixed glioma
2.2	Astrocytomas
2.3	Ependymomas
2.4	Acoustic neuromas
2.5	Craniopharyngiomas
2.6	Haemangioblastomas
2.7	Primary cerebral lymphoma
2.8	Meningiomas
2.9	Germ cell tumours
2.10	Pineal region tumours
2.11	Pituitary tumours
2.12	Primitive neuroectodermal tumours (PNETs)
2.13	Spinal cord tumours
3	Breast Cancer
3.1	DCIS- Ductal carcinoma in situ
3.2	LCIS- Lobular carcinoma in situ
3.4	Invasive ductal breast cancer
3.5	Invasive lobular breast cancer
3.6	Inflammatory breast cancer
3.7	Paget's disease
3.8	Rare types of breast cancer
3.9	Breast Cancer in men
4	Bowel Cancer
4.1	Adenocarcinoma
4.1.1	Colorectal Cancer
4.1.2	Colon Cancer
4.2	Squamous cell cancers of bowel
4.3	Carcinoid tumours
4.5	Sarcomas of colon
4.6	Cancers of the lymphatic system in colon
5	Anal Cancer
5.1	Epidermoid cancers.
5.1.1	Large cell keratinising
5.1.2	Large cell non keratinising (also called transitional)
5.1.3	Basaloid
5.2	Non Epidermoid cancer
5.3	Adenocarcinoma of glandular cells
5.4	Basal cell carcinoma of anal origin
5.5	Melanoma of anal origin

5.6	Rectal cancer or cancer of the rectum
6	**Womb Cancer**
6.1	Endometrial cancer
6.1.1	Endometrioid adenocarcinomas
6.1.2	Papillary serous carcinomas
6.1.3	Clear cell carcinoma
6.2	Adenocarcinoma with squamous cells of womb
6.3	Sarcoma of the womb
6.4	Cancer of the neck of the womb (cervix)
7	**Cervical Cancer**
7.1	Squamous cell cancer of cervical
7.2	Adenocarcinoma of cervical
8	**Ocular cancer or Eye cancer**
8.1	Melanoma of the eye
8.2	Lymphoma of the eye
8.3	Rare cancers in children
9	**Gall bladder cancer**
9.1	Adenocarcinoma Gallbladder
9.2	Squamous cell cancer of the gallbladder
9.3	Adenosquamous cancer of the gallbladder
9.4	Small cell cancer of the gallbladder
9.5	Sarcoma of the gallbladder
9.6	Neuroendocrine tumour of the gallbladder
9.7	Lymphoma and melanoma of the gallbladder
10	**Stomach Cancer**
10.1	Adenocarcinoma of the stomach
10.2	Squamous cell cancers of stomach
10.3	Lymphoma of the stomach
10.4	Gastrointestinal stromal tumour (GIST)
10.5	Neuroendocrine tumours
11	**Kidney cancer**
11.1	Renal cell cancer
11.1.1	Clear cell
11.1.2	Papillary (Type 1 and 2)
11.1.3	Chromophobe
11.1.4	Oncocytic
11.1.5	Collecting duct
11.2	Transitional cell cancer (TCC) of the renal pelvis
12	**Larynx or Laryngeal cancer**
12.1	Squamous cell cancer of larynx
12.2	Adenocarcinoma of larynx
12.3	Sarcoma of larynx
12.4	Myeloma of larynx
13	**Liver cancer**
13.1	Hepatocellular carcinoma (HCC)
13.2	Cholangiocarcinoma
13.3	Angiosarcoma
13.4	Hepatoblastoma
14	**Lung cancer**
14.1	Small cell lung cancer
14.2	Non small cell lung cancer
14.2.1	Squamous cell carcinoma of lung
14.2.2	Adenocarcinoma of lung
14.2.3	Large cell carcinoma of Lung
14.3	Mesothelioma
15	**Lymphoma**
15.1	Hodgkin Lymphoma
15.1.1	Classical types
15.1.1.1	Nodular sclerosing
15.1.1.2	Mixed cellularity
15.1.1.3	Lymphocyte rich
15.1.1.4	Lymphocyte depleted
15.1.2	Nodular lymphocyte predominant type
15.2	Non Hodgkin lymphoma (NHL)
15.2.1	Cutaneous T cell lymphoma
15.2.2	Mucosa associated lymphoid tissue (MALT) lymphoma
15.2.3	Mantle cell lymphoma

16	**Melanoma skin cancer**
16.1	Superficial spreading melanoma
16.2	Nodular melanoma
16.3	Lentigo maligna melanoma
16.4	Acral lentiginous melanoma
16.5	Other types of melanoma
16.5.1	Internal organs melanoma
16.5.2	Melanoma of the eye
17	**Mouth and oropharyngeal cancer**
17.1	Mouth and oropharyngeal cells
17.2	Squamous cell cancers of the mouth and oropharynx
17.3	Other types of mouth and oropharyngeal cancer
17.3.1	Salivary gland cancer
17.3.1.1	Adenoid cystic cancer
17.3.1.2	Lymphoma
17.3.1.3	Melanoma
18	**Myeloma**
18.1	Light chain myeloma
18.2	Non secretory myeloma
19	**Oesophageal cancer**
19.1	Squamous cell carcinoma of oesophageal
19.2	Adenocarcinoma oesophageal
19.3	Undifferentiated cancer
19.4	Rare types of oesophageal cancer
20	**Ovarian cancer**
20.1	Epithelial ovarian cancer
20.2	Germ cell and other rare ovarian tumours
21	**Pancreatic cancer**
21.1	Exocrine cancers
21.1.1	Cystic tumours
21.1.2	Cancer of the acinar cells
21.1.3	Sarcomas of the pancreas
21.2	Endocrine pancreatic tumours
21.3	Lymphoma of the pancreas
22	**Penile (penis) cancer**
22.1	Squamous cell cancer of the penis
22.2	Adenocarcinoma
22.3	Melanoma of the penis
22.4	Basal cell cancer of the penis
22.5	Sarcoma of the penis
23	**Salivary gland cancer**
23.1	Cells of the salivary gland
23.2	Benign tumours
23.3	Mucoepidermoid cancer
23.4	Acinic cell carcinoma
23.5	Adenoid cystic cancers
23.6	Adenocarcinoma
23.7	Malignant mixed cancers
23.8	Low grade polymorphous cancers
23.9	Other types
23.9.1	Squamous cell carcinoma
23.9.2	Lymphoepithelioma
23.9.3	Anaplastic carcinoma
23.9.4	Poorly differentiated carcinoma
23.9.5	And other even rarer cancers
24	**Skin cancer (non melanoma)**
24.1	Basal cell skin cancer
24.1.1	Nodular
24.1.2	Superficial
24.1.3	Morphoeic
24.1.4	Pigmented
24.2	Squamous cell skin cancer
24.3	Rarer types of non melanoma skin cancer
24.3.1	Merkel cell carcinoma
24.3.2	Kaposi's sarcoma
24.3.3	T cell lymphoma of the skin
25	**Soft tissue sarcomas**

25.1	Fibrosarcomas
25.2	Leiomyosarcomas and rhabdomyosarcomas
25.3	Liposarcomas
25.4	Synovial sarcomas
25.5	Angiosarcomas
25.6	Malignant peripheral nerve sheath tumours
25.7	Gastrointestinal Stromal Tumours (GISTs)
25.8	Kaposi's sarcomas
25.9	Ewings sarcomas and PNETs
25.10	Fibromatosis
26	**Testicular cancer**
26.1	Seminomas
26.2	Non seminomas
27	**Thyroid cancer**
27.1	Papillary thyroid cancer
27.2	Follicular thyroid cancer
27.3	Medullary thyroid cancer
27.4	Anaplastic thyroid cancer
27.5	Rare types of thyroid cancer
28	**Vaginal cancer**
28.1	Squamous cell carcinoma of vaginal
28.2	Adenocarcinoma of the vagina
28.2.1	Clear cell adenocarcinoma
28.2.2	Papillary adenocarcinoma
28.2.3	Mucinous adenocarcinoma
28.2.4	Adenosquamous carcinoma
28.3	Sarcoma of the vagina
28.4	Melanoma of the vagina
28.5	Small cell vaginal cancer
29	**Vulval cancer**
29.1	Squamous cell carcinoma of vulval cancer
29.2	Vulval melanoma
29.3	Adenocarcinoma
29.4	Basal cell carcinoma of vulval
29.5	Verrucous carcinoma of vulval
29.6	Sarcomas of vulval
30	**Bone cancer**
30.1	Osteosarcoma
30.2	Ewing's sarcoma
20.3	Chondrosarcoma
30.4	Spindle cell sarcoma
30.5	Chordoma
31	**Bladder cancer**
31.1	Transitional cell (Urothelial cancer)
31.2	Non Muscle Invasive
31.3	Squamous cell bladder cancer
31.4	Invasive bladder cancer
31.5	Adenocarcinoma of the bladder

The following three tables, 2.6, 2.7 and 2.8 are the example tables related to section 2.3 Data preparation part.

Table 2.6: Example of two cancers entries based on one Mutation class.

HGMD ID	Genes	Locations	Cancer	Mutations Types	Number of mutations in that gene	Number of Specific mutation	Mutation Descriptions	amino acid changes	Ref
CM103943	MLH1	3p21.3	Breast & Colorectal_Cancer	Missense/Nonsense	875	285	cCGG-TGG	Arg389Trp	Jensen (2010) Breast Cancer Res Treat 120, 777

Table 2.7: Example of Split for the two cancers entries using same Mutation HGMD ID.

HGMD ID	Genes	Locations	Cancer	Mutations Types	Number of mutations in that gene	Number of Specific mutation	Mutation Descriptions	amino acid changes	Ref
CM103943	MLH1	3p21.3	Colorectal_Cancer	Missense/Nonsense	875	285	cCGG-TGG	Arg389Trp	Jensen (2010) Breast Cancer Res Treat 120, 777
CM103943	MLH1	3p21.3	Breast_Cancer	Missense/Nonsense	875	285	cCGG-TGG	Arg389Trp	Jensen (2010) Breast Cancer Res Treat 120, 777

Table 2.8: Example for the filtered data table layout, i.e. an example for one cancer only, “Bladder cancer”, where the phenotype is the cancer name, gene is a gene symbols, Mutation classes, total number of mutations for each gene including cancers and non-cancers, total number of mutation classes including cancers and non-cancers as well, location of the gene in the chromosome, the collected HGMD unique ID and finally a PubMed reference for each mutation and cancer records.

Phenotype	Gene	Mutation classes	Total Number of gene mutations	Total number of mutation class for	Chromosome location	HGMD Accession	References
Bladder Cancer	ARL6IP5	Regulatory	1	1	3p14	CR057790	Wu (2005) Zhonghua Yi Xue Yi Chuan Xue Za Zhi 22, 64
Bladder Cancer	DDX20	Missense nonsense	1	1	1p21.1-p13.2	CM086198	Yang (2008) Cancer Res 68, 2530
Bladder Cancer	ERCC4	Regulatory	22	1	16p13.3-p13.11	CR102187	Wang (2010) Oncogene 29, 1920
Bladder Cancer	ERCC6	Missense nonsense	74	35	10q11.23	CM0910264	Chang (2009) Anticancer Res 29, 5121
Bladder Cancer	HRAS	Missense nonsense	18	13	11p15.5	CM035804	Johne (2003) Cancer Epidemiol Biomarkers Prev 12,68
Bladder Cancer	MDM2	Regulatory	3	3	12q14.3-q15	CR082025	Wang (2008) Clin Cancer Res 14, 3633
Bladder Cancer	OGG1	Regulatory	17	4	3p26.2	CR072322	Figueroa (2007) Hum Genet 121, 233
Bladder Cancer	PTGS2	Regulatory	8	5	1q25.2-q25.3	CR050013	Kang (2005) Cancer Lett 217, 11
Bladder Cancer	TNFRSF10A	Regulatory	3	1	8p21	CR117979	Wang (2009) Mutat Res 661, 85
Bladder Cancer	UGT2B7	Missense nonsense	7	2	4q13	CM057512	Lin (2005) Toxicol Sci 85, 502
Bladder Cancer	XPC	Missense nonsense	55	21	3p25	CM112197	Qiao (2011) Carcinogenesis 32, 516
Bladder Cancer	XPC	Missense nonsense	55	21	3p25	CM112198	Qiao (2011) Carcinogenesis 32, 516
Bladder Cancer	XPC	Regulatory	55	2	3p25	CR112196	Qiao (2011) Carcinogenesis 32, 516
Bladder Cancer	XRCC4	Splicing	4	1	5q13-q14	CS075240	Figueroa (2007) Carcinogenesis 28, 1788
Bladder Cancer	XRCC5	Repeat Variations	1	1	2q35	CE075931	Wang (2008) Mutat Res 638, 26
Bladder Cancer	ZNF350	Missense nonsense	1	1	19q13.33	CM074642	Figueroa (2007) Carcinogenesis 28, 1788
Bladder Cancer	MLH1	Small Deletion	875	195	3p21.3	CD105829	van der Post (2010) J Med Genet 47, 464
Bladder Cancer	MSH2	Splicing	860	92	2p22-p21	CS104167	van der Post (2010) J Med Genet 47, 464
Bladder Cancer	MSH2	Gross Deletions	860	190	2p22-p21	CG104168	van der Post (2010) J Med Genet 47, 464
Bladder Cancer	MSH2	Gross Deletions	860	190	2p22-p21	CG104166	van der Post (2010) J Med Genet 47, 464

2.4. Statistical Methods

2.4.1. Cohen's Kappa (κ) Coefficient

Cohen's kappa coefficient was originally proposed by (Carletta, 1996; Cohen, 1960) to quantify the inter-rater agreement or inter-annotator agreement between two individuals. This is thought to be more robust a measure than simple percent agreement calculation, since k takes into account the agreement occurring by chance. The agreement is calculated between two raters, which each classify *N* items into *C* mutually exclusive categories. The equation for quantifying $k = \frac{Pr_{(a)} - Pr_{(e)}}{1 - Pr_{(e)}}$

Where Pr(*a*) is the relative observed agreement among raters, and Pr(*e*) is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly classifying each category randomly. If the raters are in complete agreement then κ = 1. If there is no agreement among the raters other than what would be expected by chance [as defined by Pr(*e*)], κ = 0.

However Kappa's strength lies in its ability to take into account chance agreement between two or more observers. A second strength of Kappa lies in its known sampling distribution. Kappa's weakness is that it takes no account of the degree of disagreement, i.e., all disagreements are treated equally.

Weighted Kappa Statistic

The weighted kappa coefficient is a generalization of the simple kappa coefficient, using weights to quantify the relative difference between categories. The weighted kappa is useful in programming, when codes are sequential in order (Cohen, 1968).

The equation for weighted $k = 1 - \frac{\sum_{i}^{k}=1 \sum_{j}^{k}=1 w_{ij}x_{ij}}{\sum_{i}^{k}=1 \sum_{j}^{k}=1 w_{ij}m_{ij}}$

Where k = number of codes and w_{ij}, x_{ij}, and m_{ij} are elements in the weight, observed, and expected matrices, respectively. When diagonal cells contain weights of 0 and all off-diagonal cells weights of 1, this formula produces the same value of Kappa as the calculation given above.

Interclass Correlation Class (ICC)

The interclass correlation coefficient is useful for measuring the degree of dependence of observations within a group (Koch, 1983) to determine how similar are the values of some variables within the same group. Essentially, it is a measure of the within subject (group) variability. Within each group, the mean of a value can be computed, and if values are similar then the difference between the mean and a single value will be small. The ICC is particularly relevant for work with families and siblings because of the possibility of having more than two siblings within a family (group).

The hypothesis of the current project is to quantify the degree (κ) of agreement between two or more connected cancers on the basis of the shared genes. In this study, I distinguished the association degree if two cancer into four classes: “very connected”, “highly connected”, and moderate connected” and “low connected”. Two cancers are defined as very connected if k = (0.75 to 1), highly connected if k = (0.5 to 0.75), Moderate connected if k = (0.25 to 0.5) or with low connectivity of k = (<0.25).

Theoretical example

Table 2.9: Theoretical example table. Set of five genes and four cancer types. If a gene and cancer are related the number 1 will be added to the table and if they are not related the number 0 is added.

	Gene1	**Gene 2**	**Gene 3**	**Gene 4**	**Gene n**
Cancer A	1	1	1	0	0
Cancer B	1	1	1	1	0
Cancer C	0	0	1	0	1
Cancer D	1	0	1	0	1

Table 2.10: This 3 x 3 table, based on table 2.9 above, shows the scoring of cancer entries where ‘1’ represents a positive gene association and 0 represents a no gene association between Cancer A and Cancer B.

		Cancer A		**Row total**
Cancer B		1	0	
	1	3 ($C_{1,1}$)	1 ($C_{0,1}$)	4 ($C_{1,*}$)
	0	0 ($C_{0,1}$)	1 ($C_{0,0}$)	1 ($C_{0,*}$)
Column total		3 ($C_{*,1}$)	2 ($C_{*,0}$)	5 ($T_{a,b}$)

Three steps are needed to derive the value of k;

1. Calculate the observed percentage agreement, $O_{ab} = \frac{C_{1,1} + C_{0,0}}{T_{ab}} = \frac{3+1}{5} = 0.8$
2. Calculate the probability of random agreement

$$A_{ab} = \frac{C_{*,1} * C_{1,*} + C_{*,0} * C_{0,*}}{T_{ab} * T_{ab}} = \frac{3 * 4 + 2 * 1}{5 * 5} = 0.56$$

3. Calculate the degree of k, $K_{ab} = \frac{O_{ab} - A_{ab}}{1 - A_{ab}} = \frac{0.8 - 0.56}{1 - 0.56} = 0.55$

2.4.2. Density

In this study, I am interested in investigating the density of cancer-associated genes in each chromosome as this may provide some measure structural insight. The method of mass density (density) is a measure of mass contained per unit volume. Mathematically, density is a mass divided by volume $\rho = \frac{m}{V}$, where ρ signify density, m signify mass, and V signify volume (S Dole, 2009). Here, the density formula that is going to be used to calculates the density of the collected cancer genes in each chromosome, where $m$the number of cancer genes is assigned to a particular chromosome and V is the total number of the identified genes on the associated chromosome.

2.5. Conclusion

The 3 complementary data sources, (i) Genetic Association database (GAD) (Becker et al., 2004), (ii) The Cancer Gene Census Database (COSMIC) (Futreal et al., 2004), (iii) Human Gene Mutation Database (HGMD) (Stenson et al., 2009), were successfully extracted and integrated in this study. In the GAD a total of 1,908 cancer-related entries was selected and 1,324 entries selected from COSMIC. Filtering these extracted 3,232 cancer-related gene records from both databases resulted in a list of 520 genes. Based on these filtered 520 genes, a subsequent manual extraction of 15,264-germline mutation class records were obtained including; *Missense nonsense, Splicing, Regulatory, Small deletions, Small Insertions, Small Indels. Gross deletions, Gross Insertions, Complex rearrangements and Repeat variations,* from the HGMD, for cancer and non-cancer disorders.

The newly constructed database was achieved by performing a splitting process to clarify two groups of mutation records; the first consisted of a total of 8,547 cancer mutation records and the second 6,717 non-cancer mutation records. The clarified 8,547-cancer mutation records, was successfully screened further by introducing a 'hierarchy' to cancer disorder names (family tree of cancers) using the Cancer Research UK and PubMed literature. The introduced hierarchy consisted of 31 primary cancer name terms and 206 secondary and tertiary cancer names. The newly screened data was stored in a table as an independent record for each gene symbol. Each record consisted of Primary cancer name, Gene symbol, the implicated mutation class name involved in the cancer, the

total number of evident mutations for each gene including cancers and non-cancers mutations, the total number of mutations for each of the implicated class, gene chromosome location and PubMed literature reference. In total I achieved 8,879 Mutations records including 424 unique genes and 29 primary cancer disorder names.

The method Choen's Kappa Coefficient was used to measure the agreements between two cancers in order to determine if there can be connected nodes in the to-be constructed Human cancer map.

These achieved our objectives in relation to the biological data and detected mathematical methods were then used in the next two chapters to extract some biological meaning in order to increase our understanding of how mutation can impact on cancer development.

Chapter Three

3. Association of Human Cancers, Genes and Mutations Classes

Cancers can develop in response to multiple contributory factors and can be associated through various factors (principally genetic ones). Many authors have shown the value of revealing cancer associations based on cancer-causing genes (Goh et al., 2007; Barabási et al., 2011). Others have assessed cancer development based on an analysis of gene expression to identify the proteins that may act on similar disease-causing pathway(s) (Dreze et al., 2010; Goehler et al., 2004). The aim of the work presented in this chapter is to investigate the associations of cancers based on the involvement or usage of particular genetic mutation classes. Using this approach, I intend to explore whether genetic mutation classes vary from one to another in terms of their associations to cancers. The curated mutation data shows that the association of various cancers correspond to gene mutation(s) very differently. For example, Breast cancer is associated to ten different types of mutations in each of several genes, while Bone cancer is associated to just two different types of

mutation based on the same causing gene. In order to gain greater insight into the nature of the associations between cancers, and the association of gene mutation type and the target cancers, I conducted a series of investigations, including: (i) constructing and analysing a Human Cancers Map (HCM) based on genetic mutations in human genome, (ii) investigation into the distribution of mutations classes over the HCM (iii) constructing a Genome-wide Distribution Map to identify the involved genes in the HCM (IV) identifying pathways for the involved target genes, (V) to cluster genes based on the relatedness of their biological factors.

3.1. A Human Cancer Map Based on Genes and Their Associated Mutation Classes

The map connecting human cancers for investigating the association between cancer disorders was manipulated by running a string (loop statement) in R, to loop the 8,879 curated mutation records based on table 2.8 (see chapter 2). Two cancer disorders are connected if they involve a common gene due to same mutation class and not only with associated gene itself (Figure 3.1).

The Human Cancer Map (HCM) displayed many connections between cancer nodes, with each of these connections corresponding to an implicated gene and a contributing mutation class. Of 26 primary cancers, 20 had at least one link to other cancers in the HCM. Sixty-nine genes (from total gene set of 424) underpinned these links between cancers.

This suggests that the genetic origin and associated mutation class plays a central role in associations between cancer nodes.

Next, the degree of connectivity (k) of nodes distributed in the HCM was assessed. Seven cancers, representing one particular hub in the HCM, were found to be connected to a large number of other cancers as indicated by their high k value. For example, Bowel cancer (k = 18), Lung cancer (k =17), Breast cancer (k = 16), Brain tumours (k = 13), Ovarian cancer (k =13), Stomach cancers (k = 12), and Melanoma (k = 11). This finding indicates that each of these cancers is associated with several other cancers via the same gene and mutation classes (Figure 3.1).

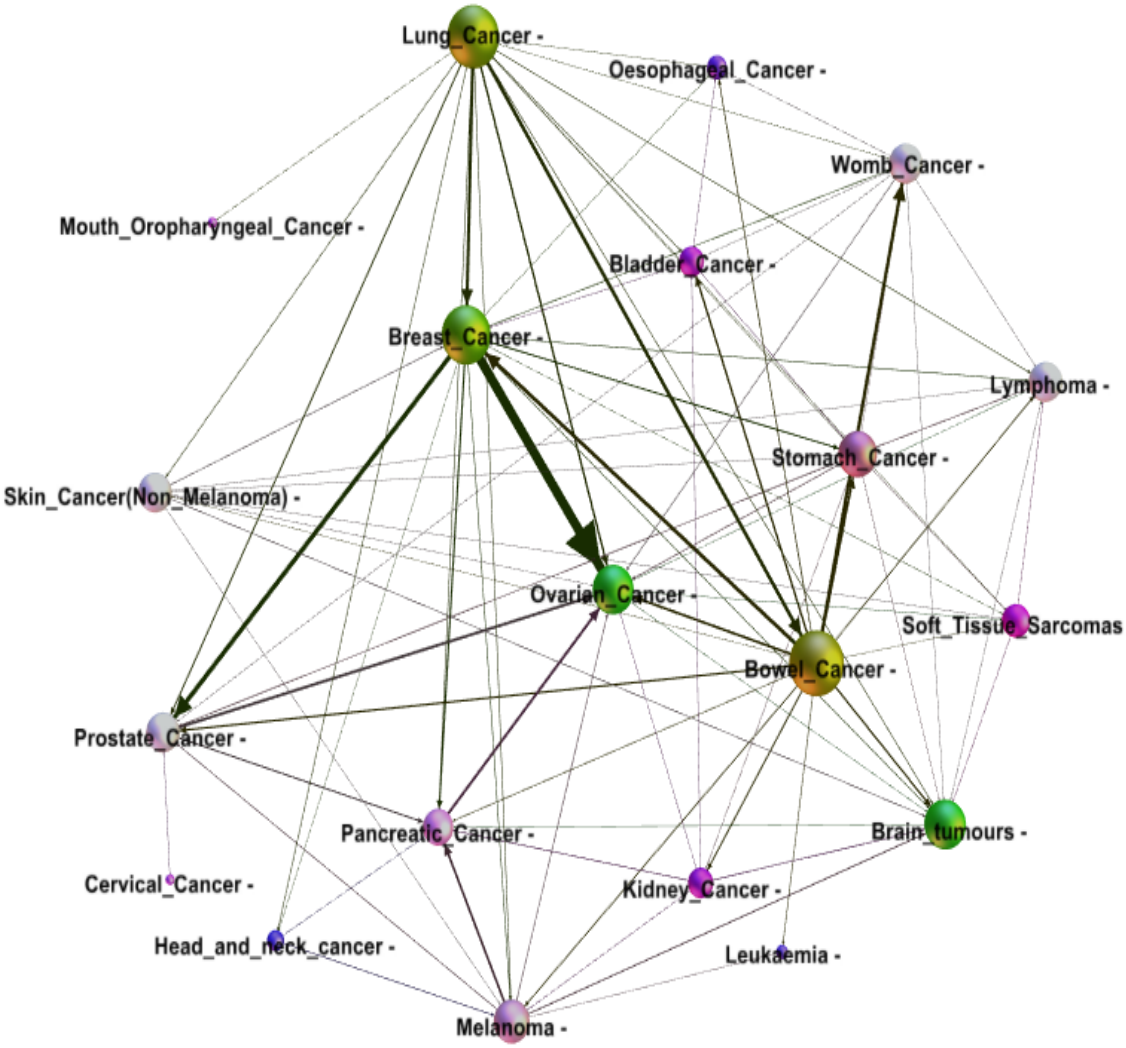


Figure 3.1: Human Cancer Map (HCM). Each node corresponds to a primary cancer and a link between two nodes represents the presence of shared cancer genes and mutation class.

Key to Figures 3.1 to 3.12. The size of each node is proportional to the number of interconnections shared by that node i.e., number of distinct primary cancers connecting to it. The colours assigned to the node represent the number of interconnected links between primary cancer types, and the width of the link-lines reflects the number of genes linking each cancer nodes.

3.1.1. Investigations into the Distribution of Mutation Classes in a Constructed Human Cancer Map

The distribution of mutation classes behind the Human Cancer Map (HCM) was next investigated, to gain a clearer understanding of the impact of the underlying mutation class to various cancers. This investigation was based on extracting a mutation-based classes-HCM (CM-HCM) from the main HCM (Figure 3.1), with each map corresponding to a specific mutation type (see figures 3.2~3.11) and the corresponding tables (A1.1~A1.10).

The ratio for the distribution of each mutation class was calculated based on the total number of the interconnections (links) between cancers nodes for each of the CM-HCMs divided by the total number of interconnections for the whole HCM. In total there are 171 interconnections in the HCM. Of these 44% were for Missense/Nonsense; 14% small deletion, 9% splicing, 10% regulatory, 10% small insertion, 6% gross deletion, 3% small indels, 1% gross insertion, 1% complex rearrangement and finally 2% repeat variation (Figure 3.12).

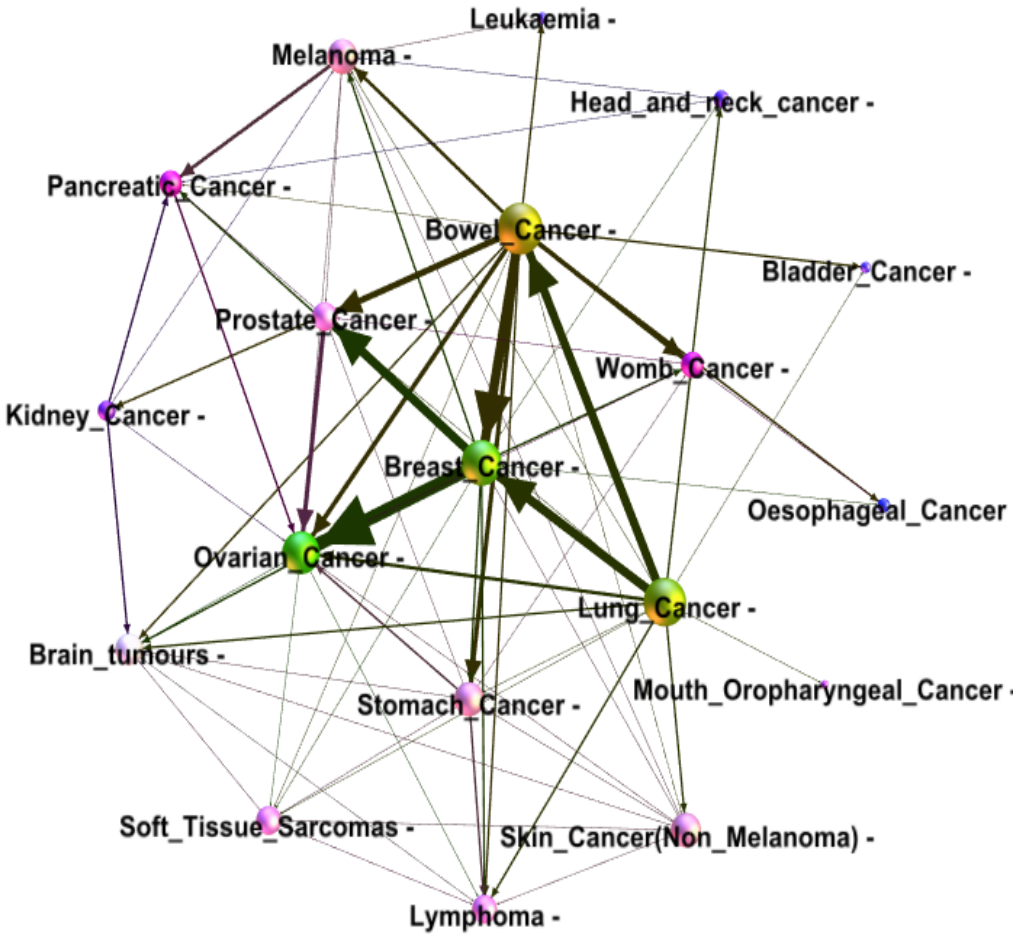


Figure 3.2: Missense/Nonsense CM-HCM. Each node corresponds to a primary cancer with a total of 19 notes and 76 links. The size of the node is proportional to the number of links. Two cancer nodes are connected if there is at least one identical gene implicated in both nodes. The width of the links (connections) is proportional to the number of genes implicated in both cancer nodes. For a full list of cancer nodes and associated genes please see Appendix 1 table A1.1.

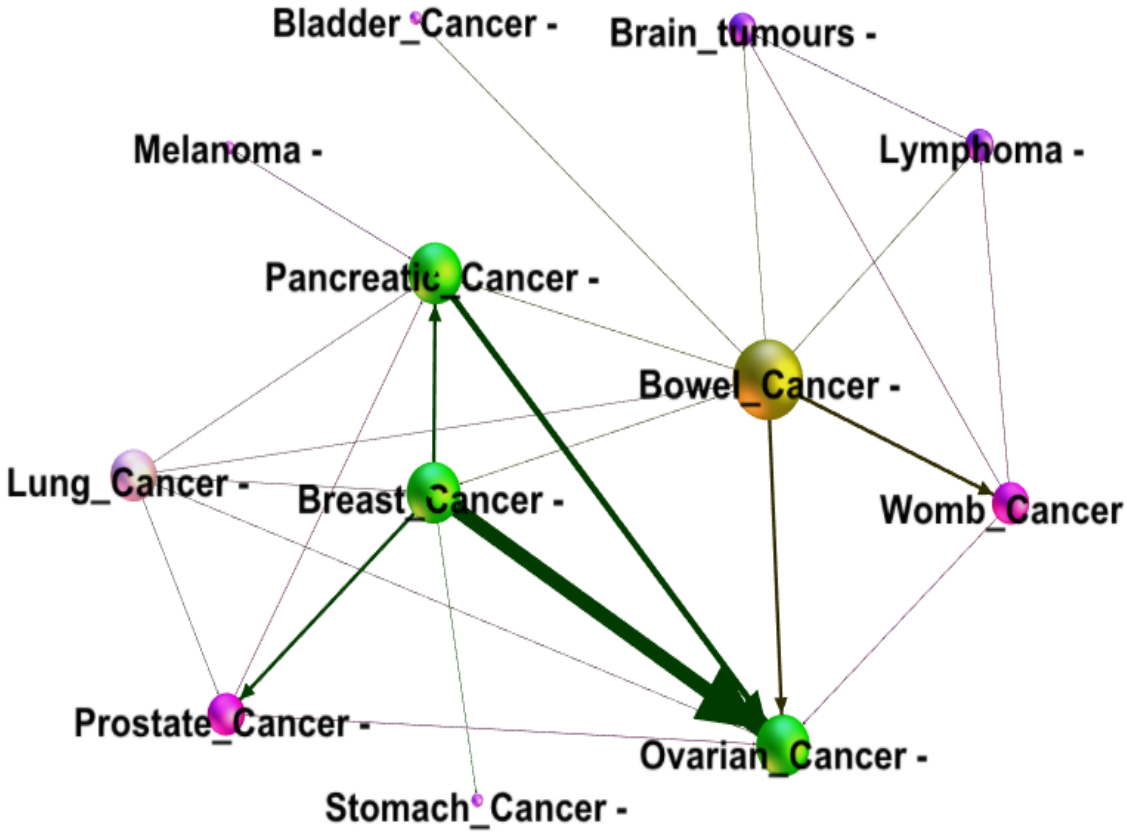


Figure 3.3: Small deletion CM-HCM. Each node corresponds to a primary cancer, totalling 12 nodes with 24 links. Full lists of cancer nodes and associated genes please see Appendix 1 table A1.2.

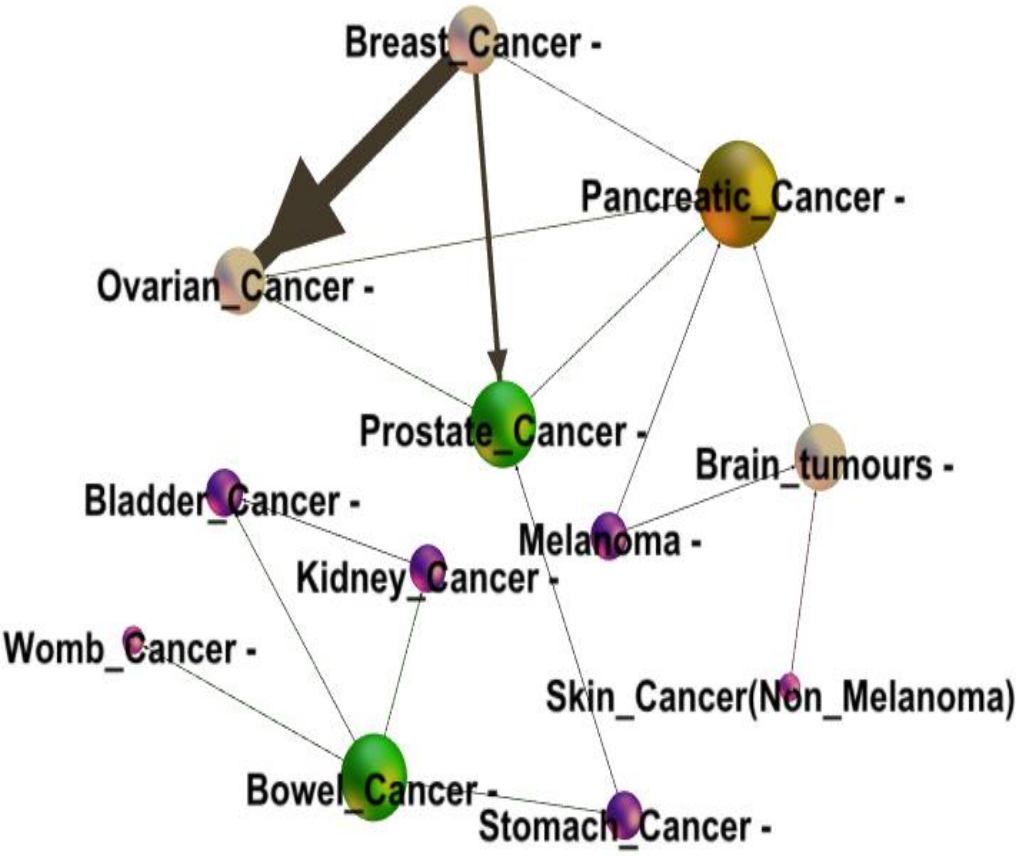


Figure 3.4: Splicing CM-HCM. Each node corresponds to a primary cancer in which there are a total of 12 nodes with 16 links. For full lists of cancer nodes and associated genes please see Appendix 1 table A1.3.

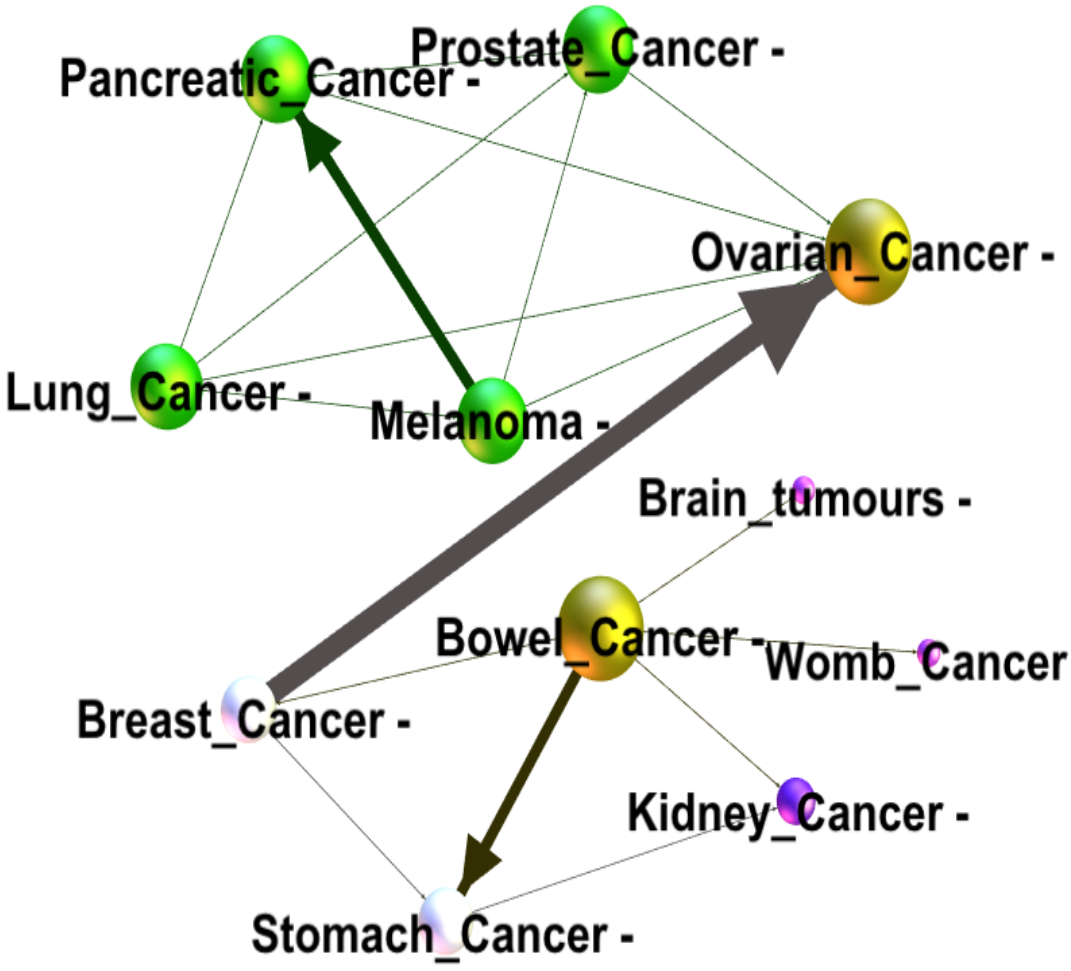


Figure 3.5: Small Insertions CM-HCM. Each node corresponds to a primary cancer and a total of 11 nodes with were detected for this SM-HCM. 18 links For full lists of cancer nodes and associated genes please see Appendix 1 table A1.4.

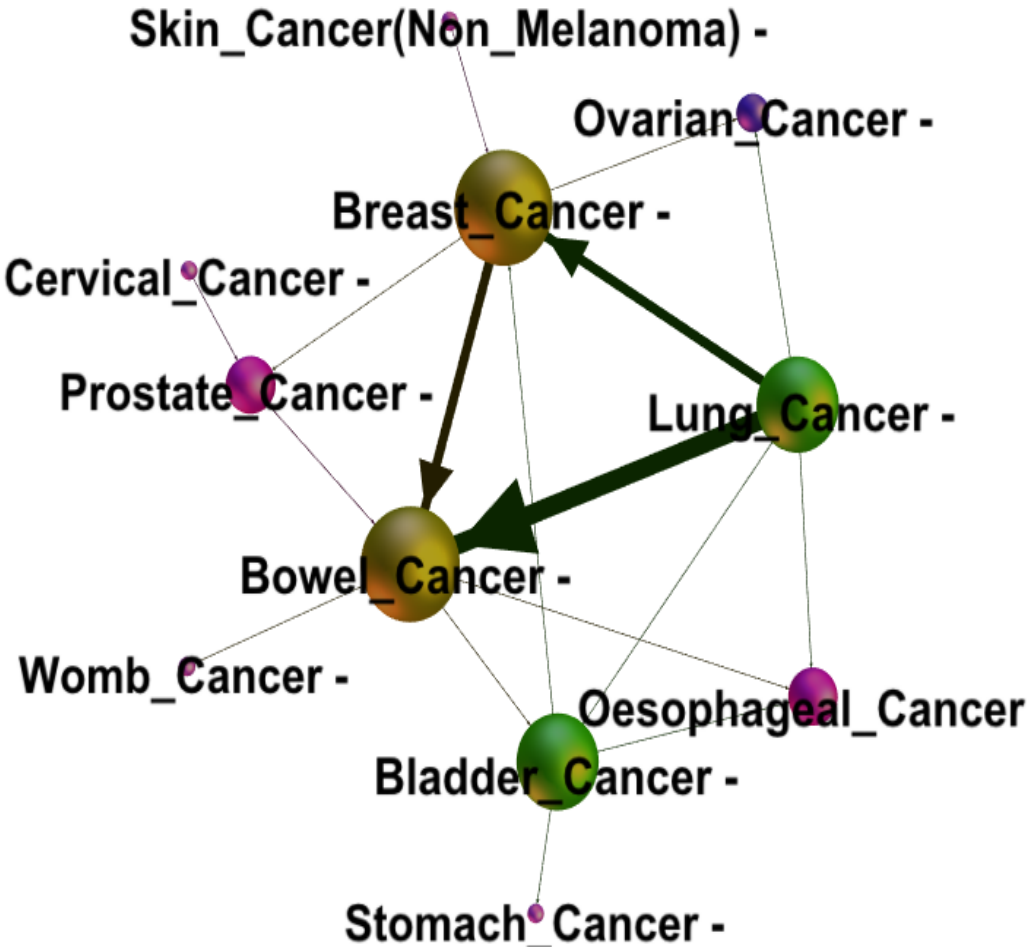


Figure 3.6: Regulatory CM-HCM. Each node corresponds to a primary cancer, with a total of 11 nodes with; 17 links for the regulatory SM-HCM network. For a full list of cancer nodes and associated genes please see Appendix 1 table A1.5.

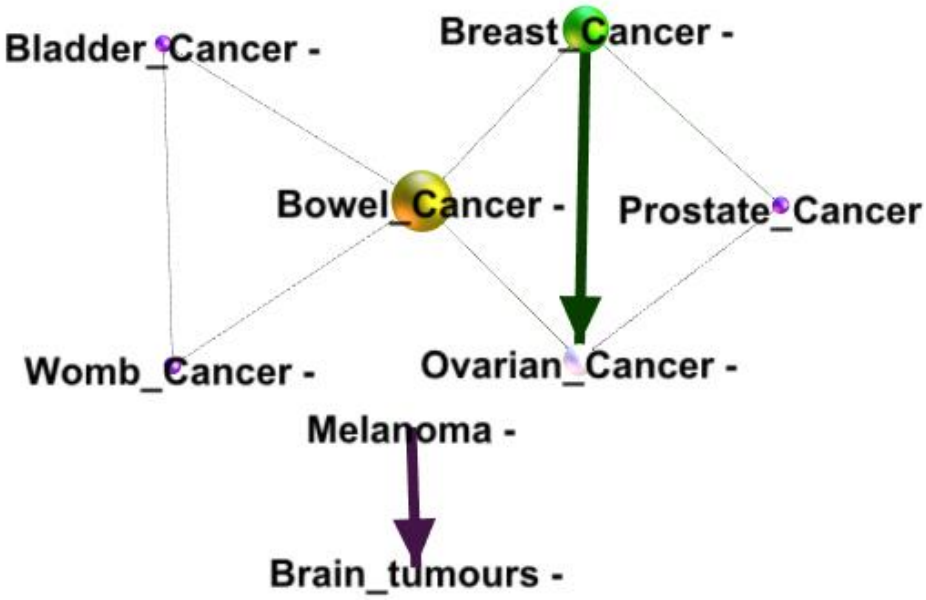


Figure 3.7: Gross deletion CM-HCM. Each node corresponds to a primary cancer, with 8 nodes and a total of 10 links for this SM-HCM. For a full list of cancer nodes and associated genes please see Appendix 1 table A1.6.

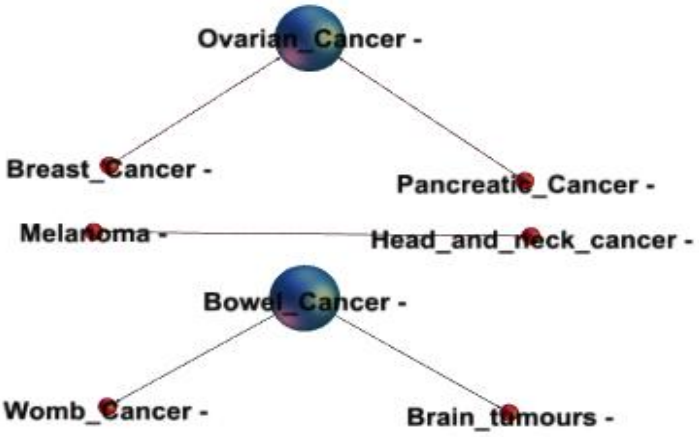


Figure 3.8: Small Indels CM-HCM. Each node corresponds to a primary cancer, with a total of 8 nodes with, 5 links for this small indels network. For a full list of cancer nodes and associated genes please see Appendix 1 table A1.7.

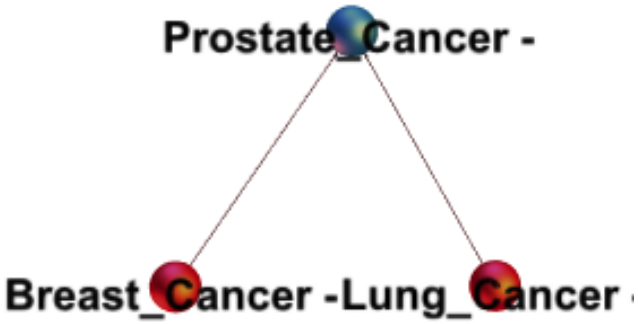


Figure 3.9: Repeat Variation CM-HCM. Each node corresponds to a primary cancer with a total of 3 nodes and 3 links for this, gross deletion network. For a full list of cancer nodes and associated genes please refer to Appendix 1 table A1.8.

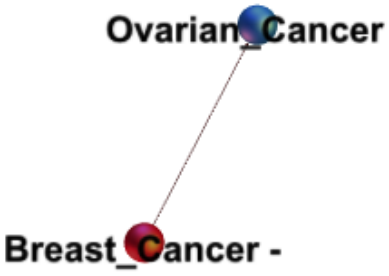


Figure 3.10: Complex Rearrangement HCM. Each node corresponds to a primary cancer in which there is a total of 2 nodes with a single link. See Appendix 1 table A1.9.or a full list of cancer nodes and associated genes.

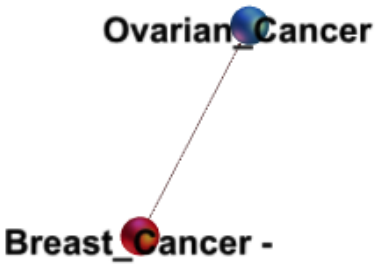


Figure 3.11: Gross Insertion HCM. Each node corresponds to a primary cancer, with a total of 2 nodes with a single link. For a full list of cancer nodes and associated genes please refer to Appendix 1 table A1.10.

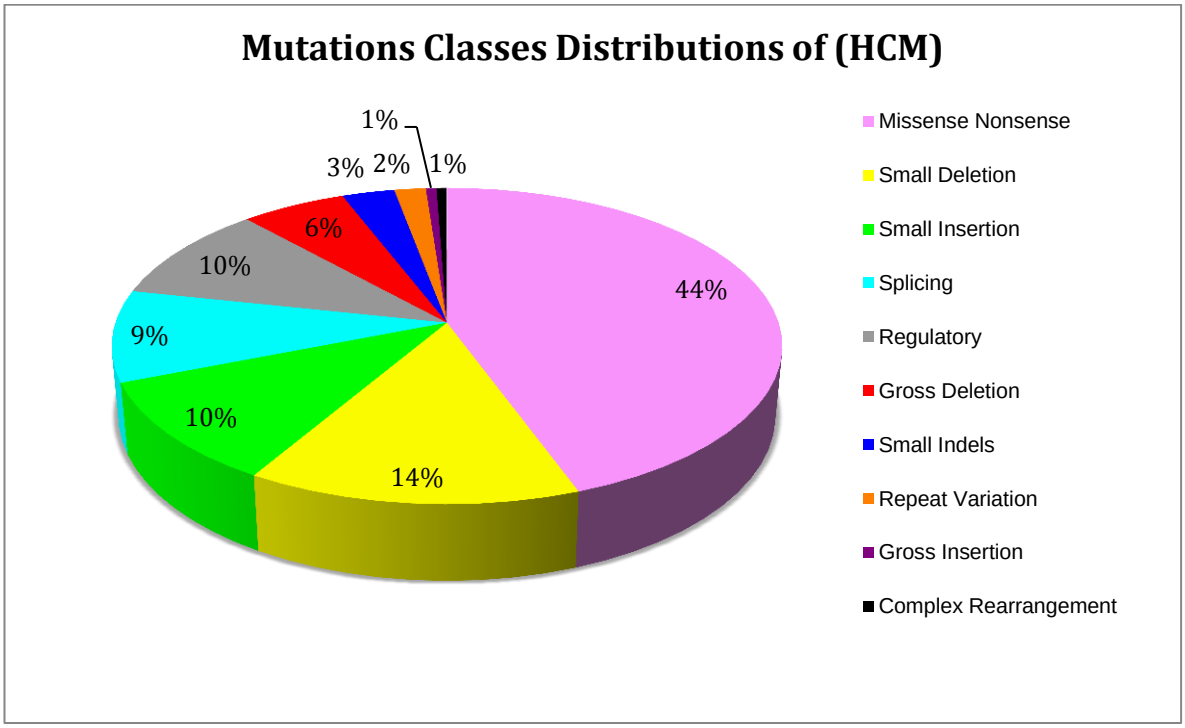


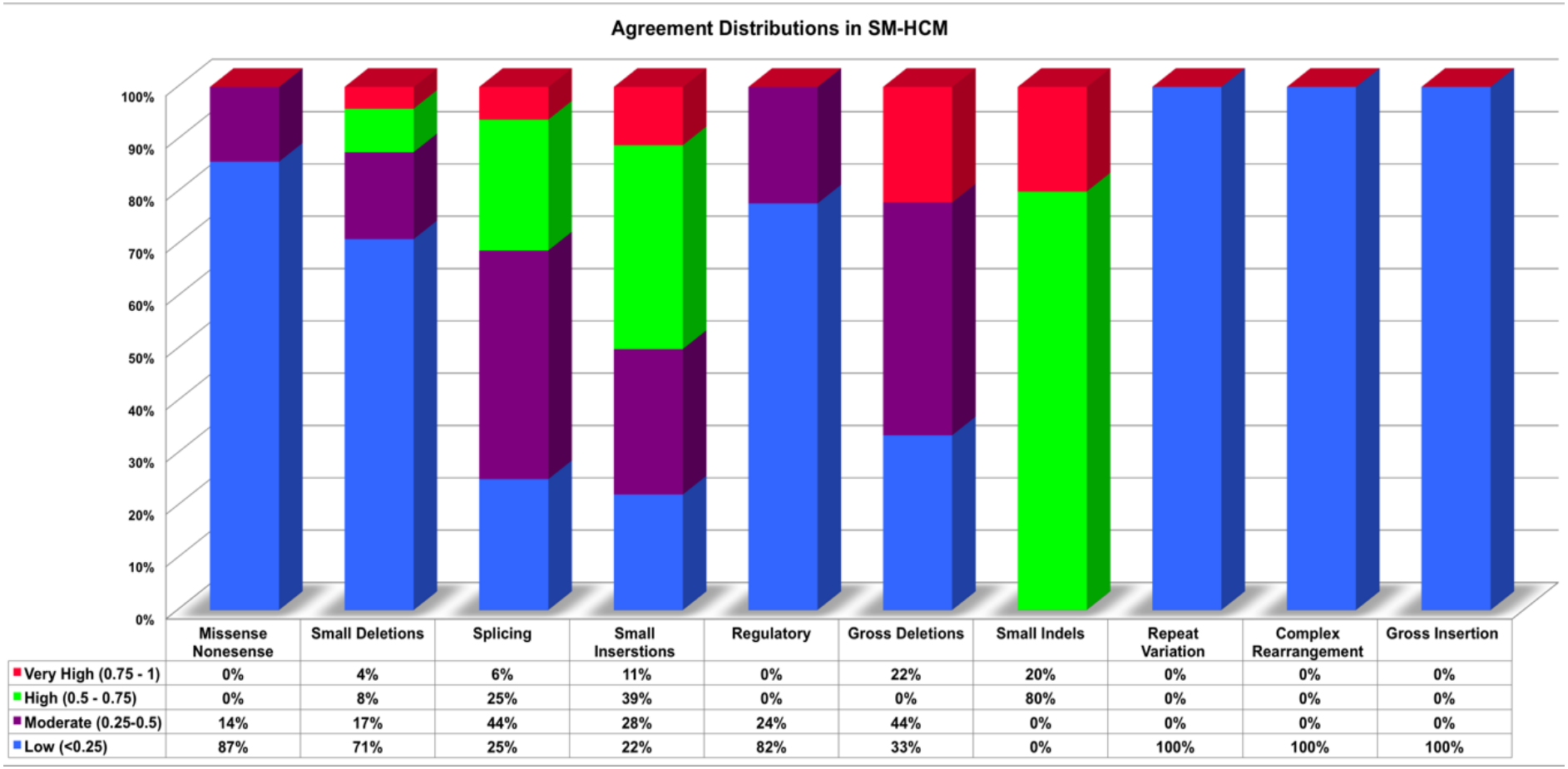
Figure 3.12: Distribution of mutation classes. The pie chart shows the % distribution of interconnections for each of the 10 mutation subclasses within the HCM.

3.1.2. The Extent of Agreement between Cancers interconnected Nodes Classes Mutation- Human Cancer Map.

The method of Choen's Kappa was applied to quantify the agreement between two connected cancer nodes in the CM-HCMs. The agreement revealed a number of interesting relationships between cancer nodes across the SM-HCMs (Figure 3.13): (1) Missense/Nonsense mutations, which formed 44% of the CM-HCM and regulatory mutation that formed 10% of the CM-HCM both showed low to moderate agreements between their connected nodes; (2) Small deletions, splicing, small insertions, gross deletions and small indels, which formed between 3% to 14% of the CM-HCM exhibited low to very high agreements. (3) Repeat variation, complex rearrangement and cross insertion mutations, which formed only 1% to 2% of the CM-HCM showed very low agreements between their interconnected cancer nodes, Refer to appendix 1 tables (A1.1, A1.2, A1.3, A1.4, A1.5, A1.6, A1.7, A1.8, A1.9 and A1.10) for the full cancer node agreements details.

Subsequently, two of the connected cancer nodes i.e., breast and ovarian cancers, appeared to be highly agreed for splicing, small insertions and small indels mutations for the following six genes (*BRCA1, BRCA2, BRIP1, RAD51C, RAD51D, TP53*). This finding confirms the results of several previous studies (Welcsh and King, 2001; Grzybowska et al., 2002), showing that germline mutations predisposes to breast and ovarian cancer in the *BRCA2* and *BRCA1* genes. Similarly, the current study found that brain tumours tend to be highly linked to melanoma, womb and bowel

cancer in terms of their associated splicing, small insertion, gross deletion and small indels mutations for three genes (*MLH1, CDKN2A* and *BRAC2*). On the other hand, it was evident that melanoma and pancreatic cancer nodes were highly agreed in terms of their associated splicing and small insertion mutations for two genes (*CDKN2A, BRAC2*). This is supported by several scientific reports (Whelan et al., 1995; Goldstein, 2004; de Snoo et al., 2008) showing strong associations between pancreatic and melanoma. The current study also revealed that melanoma exhibited very high agreement with the head and neck cancer at the small indels mutations by the *CDKN2A* gene, which have been shown by (Cabanillas et al., 2011) in their study of a novel germline mutation in the *CDKN2A* gene in head and neck squamous cell carcinomas and melanoma. The gene *CDKN2A* is therefore an active gene as its mutations tends affact many cancers. Similarly, prostate cancers agreed too many other cancers such as melanoma, pancreatic cancer and lung cancer in terms of their shared small insertion mutation in the *BRAC2* gene.



	Missense Nonesense	Small Deletions	Splicing	Small Inserstions	Regulatory	Gross Deletions	Small Indels	Repeat Variation	Complex Rearrangement	Gross Insertion
Very High (0.75 - 1)	0%	4%	6%	11%	0%	22%	20%	0%	0%	0%
High (0.5 - 0.75)	0%	8%	25%	39%	0%	0%	80%	0%	0%	0%
Moderate (0.25-0.5)	14%	17%	44%	28%	24%	44%	0%	0%	0%	0%
Low (<0.25)	87%	71%	25%	22%	82%	33%	0%	100%	100%	100%

Figure 3.13: Agreement Distributions in CM-HCM. Blue shows low agreements (<0.25), purple represents moderate agreements ((0.25 to 0.50), green reflects all high agreement (0.50 to 0.75), while red refers to the very high agreement (0.75 – 1). The mutations subclasses names are placed under each bar.

3.2. A Genome-Wide Distribution Map for Cancer Genes

A Genome-wide Distribution Map for Cancer Genes was constructed, based on the data described in table 2.8 of chapter 2, in order to explore the association between the identified 424 cancer genes and the 26 primary cancers mutation class and distributions as well as their positions in the human genome. In order to probe the genome for these 424 genes the start and end position of the corresponding genes in the human genome are needed. One of the most comprehensive genetic sequence databases, i.e. GeneCards, (Rebhan et al., 1997; Safran et al., 2002b) was used to extract the start and end position of each gene in genome. The Genome-wide Map was constructed using CIRCOS visualizing software package, which was used to construct a distribution of Genome-wide mutations in humans (Figure 3.14).

Subsequently, the density is mass divided by volume $p = \frac{m}{V}$, where ρ signify density, m signify mass, and V signify volume. Here the density of the explored genes was calculated for each chromosome as well as the distributions of the ten mutation classes explored in this study (see figure 3.15).

The following interesting characteristics were revealed: (1) Chromosome 17 is a highly represented chromosome for cancer-associated genes, whereas chromosome X is the lowest chromosome for adopting cancer-associated genes, (2) Chromosomes 2,3,7,10,17 and 22 all contained genes represented in all the 10 different mutation classes, (3) Missense nonsense mutations were a highly spread mutation class over the entire

human genome, whereas the complex rearrangement mutation class exhibited the lowest spread of mutations over the human genome (see figure 3.15). Please note, these observations apply only to the 424 gene set used here that relates to 26 primary cancers (includes the top 10 most common cancers to affect humans worldwide) and 10 mutation classes. However, these data cannot be used to infer relevance to other cancers and other disorders. For statically significance evaluation, subsequent studies will need to interrogate clinical data and include more genes, cancers and mutation recorders to gain a fully comprehensive view of the impact of both mutation class and individual chromosome involvement. This was not within the scope of this work.

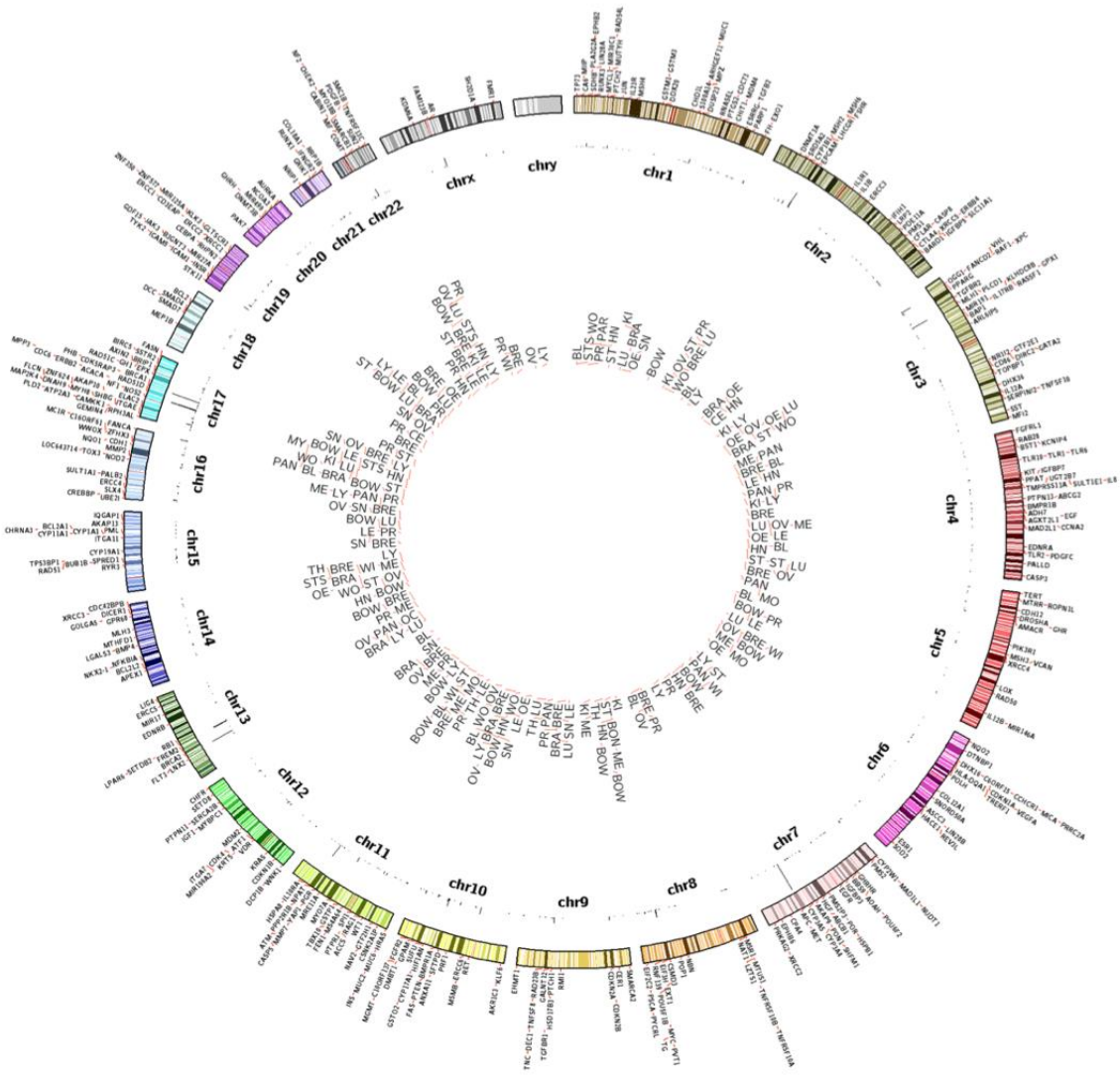
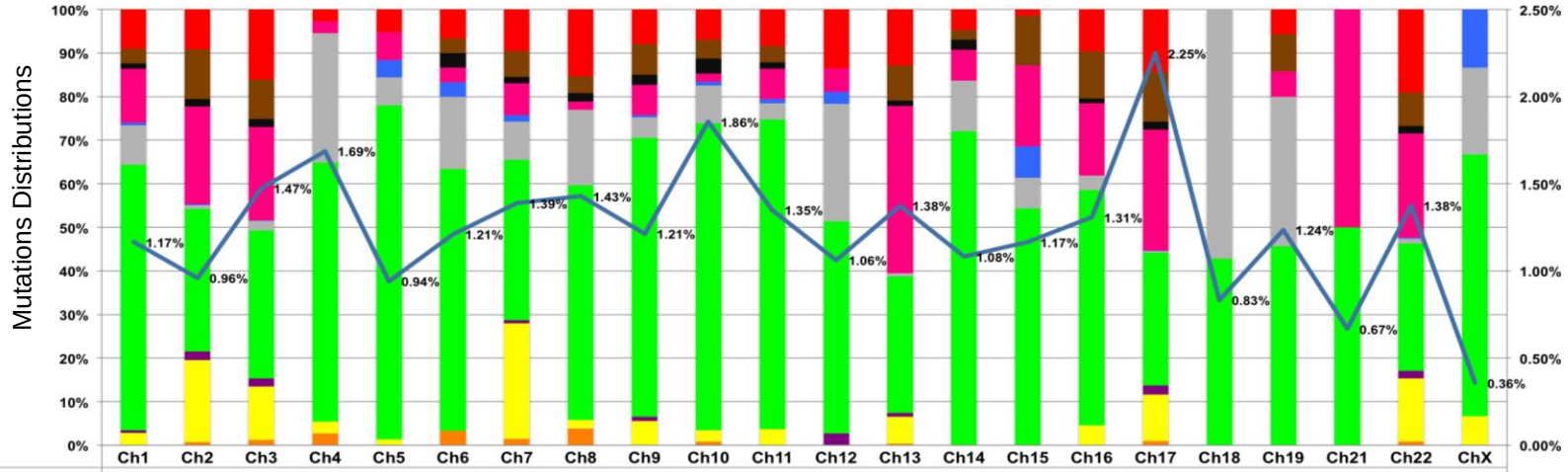


Figure 3.14: Genome-wide Distribution map for cancer genes. Blocks in the outer circle, to which are attached the 424 cancer genes, reflect chromosomes. Between chromosomes blocks and chromosome number are black bars representing the total number of mutations for each gene. The inner circle contains the abbreviations for cancers names for each chromosome. For full details of genes, cancers and mutations see table A.2.1 Appendix 2.



	Ch1	Ch2	Ch3	Ch4	Ch5	Ch6	Ch7	Ch8	Ch9	Ch10	Ch11	Ch12	Ch13	Ch14	Ch15	Ch16	Ch17	Ch18	Ch19	Ch21	Ch22	ChX
Splicing	9.04%	9.16%	16.14%	2.70%	5.19%	6.67%	9.56%	15.38%	7.94%	6.96%	8.41%	13.51%	12.91%	4.65%	1.43%	9.66%	14.41%	0.00%	5.71%	0.00%	19.08%	0.00%
Small Insertions	3.39%	11.41%	8.94%	0.00%	0.00%	3.33%	5.88%	3.85%	7.01%	4.35%	3.74%	0.00%	8.04%	2.33%	11.43%	10.80%	11.38%	0.00%	8.57%	0.00%	7.68%	0.00%
Small Indels	1.13%	1.75%	1.85%	0.00%	0.00%	3.33%	1.47%	1.92%	2.34%	3.48%	1.40%	0.00%	1.24%	2.33%	0.00%	1.14%	1.83%	0.00%	0.00%	0.00%	1.75%	0.00%
Small Deletion	12.43%	22.40%	21.48%	2.70%	6.49%	3.33%	7.35%	1.92%	7.01%	1.74%	7.01%	5.41%	38.38%	6.98%	18.57%	16.48%	27.76%	0.00%	5.71%	50.00%	23.90%	0.00%
Repeat Variations	0.56%	0.17%	0.11%	0.00%	3.90%	3.33%	1.47%	0.00%	0.47%	0.87%	0.93%	2.70%	0.00%	0.00%	7.14%	0.00%	0.11%	0.00%	0.00%	0.00%	0.22%	13.33%
Regulatory	9.04%	0.92%	2.18%	29.73%	6.49%	16.67%	8.82%	17.31%	4.67%	8.70%	3.74%	27.03%	0.70%	11.63%	7.14%	3.41%	0.28%	57.14%	34.29%	0.00%	1.10%	20.00%
Missense Nonsense	61.02%	32.72%	33.91%	59.46%	76.62%	60.00%	36.76%	53.85%	64.02%	70.43%	71.03%	48.65%	31.38%	72.09%	54.29%	53.98%	30.46%	42.86%	45.71%	50.00%	29.17%	60.00%
Gross Insertions	0.56%	1.92%	1.85%	0.00%	0.00%	0.00%	0.74%	0.00%	0.93%	0.00%	0.00%	2.70%	0.84%	0.00%	0.00%	0.00%	2.14%	0.00%	0.00%	0.00%	1.75%	0.00%
Gross Deletions	2.82%	18.90%	12.32%	2.70%	1.30%	0.00%	26.47%	1.92%	5.61%	2.61%	3.74%	0.00%	6.21%	0.00%	0.00%	4.55%	10.61%	0.00%	0.00%	0.00%	14.47%	6.67%
Comples Rearrangements	0.00%	0.67%	1.20%	2.70%	0.00%	3.33%	1.47%	3.85%	0.00%	0.87%	0.00%	0.00%	0.30%	0.00%	0.00%	0.00%	1.02%	0.00%	0.00%	0.00%	0.88%	0.00%

Figure 3.15: Distributions of genes and genetic mutations over all human chromosomes. Here the bars represent cancer mutations in each of the 23 chromosomes. The component colours of the blocks on these bars reflect a different subclass of mutations. The blue line across the chart represents the density of cancer genes in each of the chromosomes. Written below the chart are the densities for each of the mutation subclasses.

3.2.1. Identification of Genes Involved in a Human Cancer Map, Using their Genome-Wide Distribution

Identification of genes involved in the Human Cancer Map (Figure 3.1) was performed based on the gathered underlying gene records and the mutation classes records involved in driving the connections between cancer nodes in the HCM. In this way 69 genes were discovered from 4,964 mutation records, which are involved in 20 primary cancers. Based on these 4,964 mutation recorders, a string in R was run to associate genes to each other. Two or more genes were associated if they are both involved with the same cancer via the same mutation class. For example the genes *XPC, ERCC6* are considered to be associated each other because both tended to be causative genes for Bladder cancer at the same Missense/Nonsense mutation. Meanwhile the gene *MLH1*, will be excluded from an association for this particular cancer because it is associated by a different type of mutation class (i.e., small deletions), whereas Bowel cancer can associated the three genes (*XPC, ERCC6* and *MLH1*) based on missense nonsense mutation. As a result a total of 1,158 associations/connections between the 69 cancer genes were extracted. These were further explored in CIRCOS to construct a genome-wide distribution map of human cancer, and so present the 1,158 associations between cancer genes and their positions in the human genome (Figure 3.16).

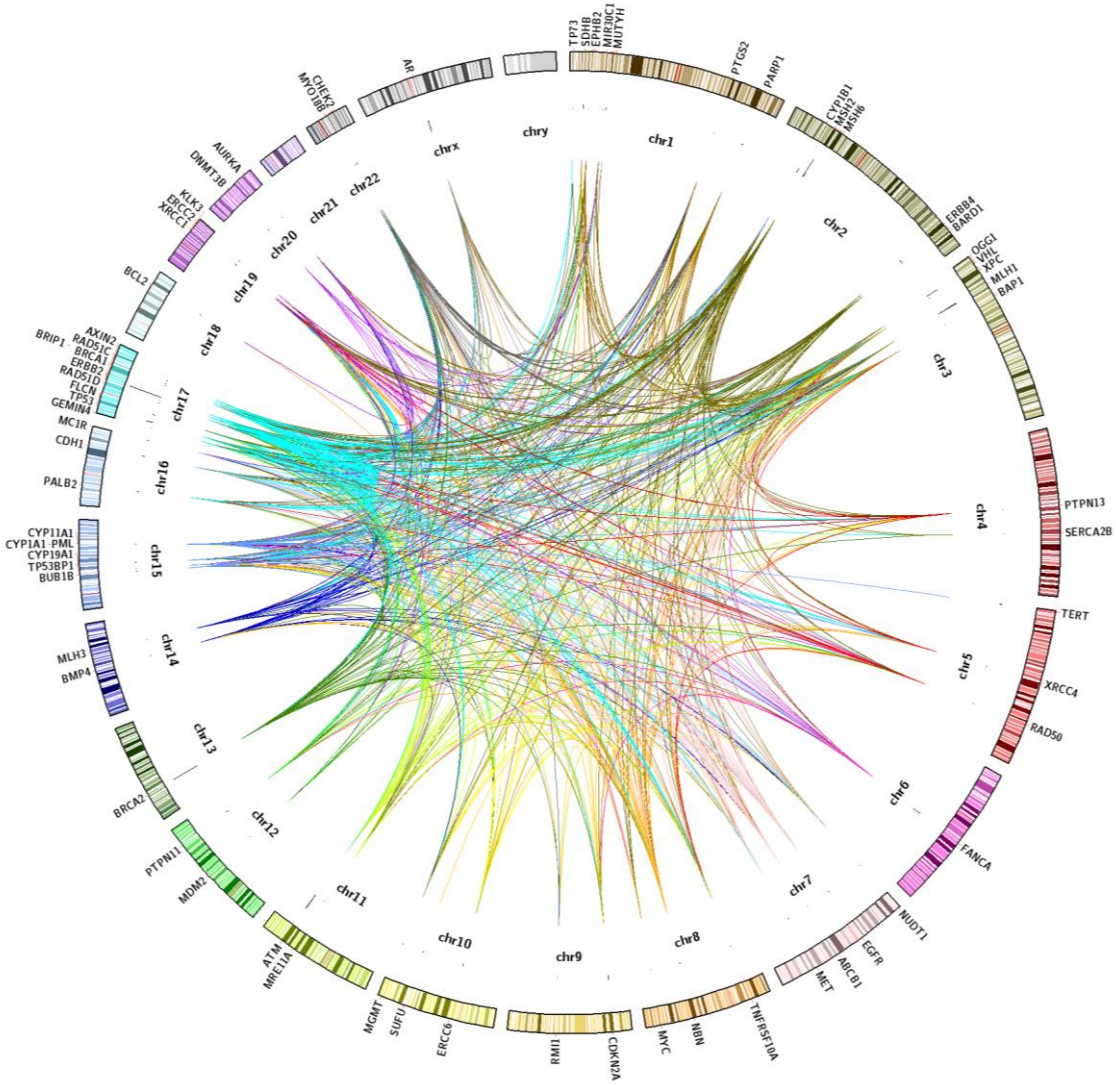


Figure 3.16: Genome-wide mutations, where the blocks in the outer circle reflect particular chromosomes and 69 genes attached to each chromosomes position. The inner interconnections represent associations between two or more genes connected if they contribute to same cancer and the same mutation class. The colours of the interconnections represent the colours of the corresponding chromosome.

3.2.2. Investigations on the Distribution of Chromosomes, Mutation Classes and Genes in a Genome-Wide Distribution Map

The investigation of the distribution of chromosomes, mutation classes and genes in the constructed genome-wide distributions (Figure 3.16) was done by extracting data of Figure 3.16 into eight genome-wide distribution figures, each of which represents a mutation subclass with interconnections of the genes cross the human genome (Figures 3.17 to 3.24). Thereafter the distributions of the 69 associated genes in each chromosome were calculated (Figure 3.25) to show the number of implicated genes on each chromosome. Ch17, Ch1, Ch15, Ch2 and Ch3 tended to contribute to cancer genes disproportionally compared with others, whereas Ch22 and Chy did not show any contribution to any of the cancer –associated genes.

As outlined in eight sub-genome-wide mutation diagrams (figures 3.17 to 3.24), it was clear that Missense/Nonsense mutations has the highest impact to human cancer as contributing broadly across the human chromosomes, as well in the interconnections between genes. This indicates that Missense/Nonsense mutations could be ‘driver mutation’ in these interconnecting genes, followed by other mutation classes (i.e., *regulatory, small deletions, splicing, gross deletions, small insertions, small indels and repeated variations*). Surprisingly gross insertion and complex rearrangement mutations had no impact on the interconnecting genes.

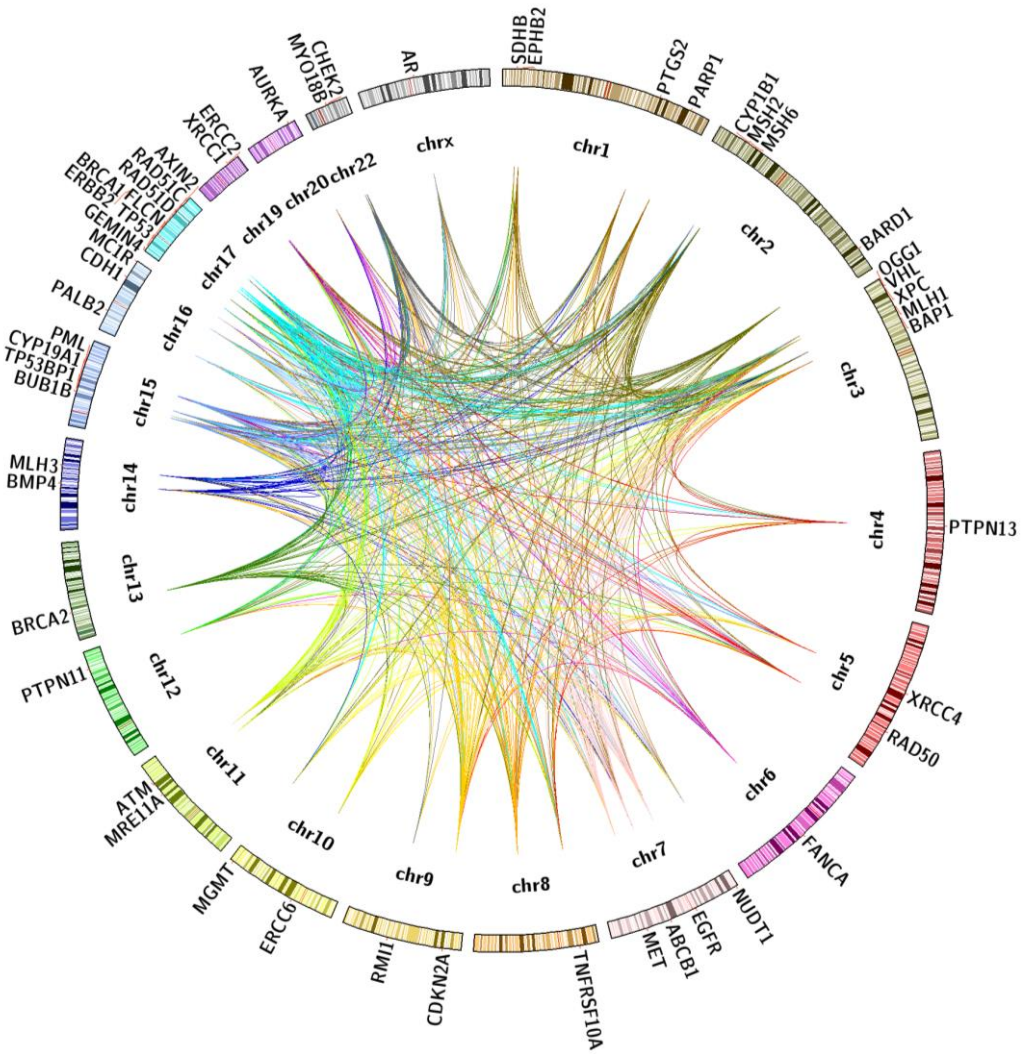


Figure 3.17: Missense/Nonsense distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes with the 54 genes attached to their corresponding chromosomes. The inner interconnections represent associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the corresponding colour of the chromosome.

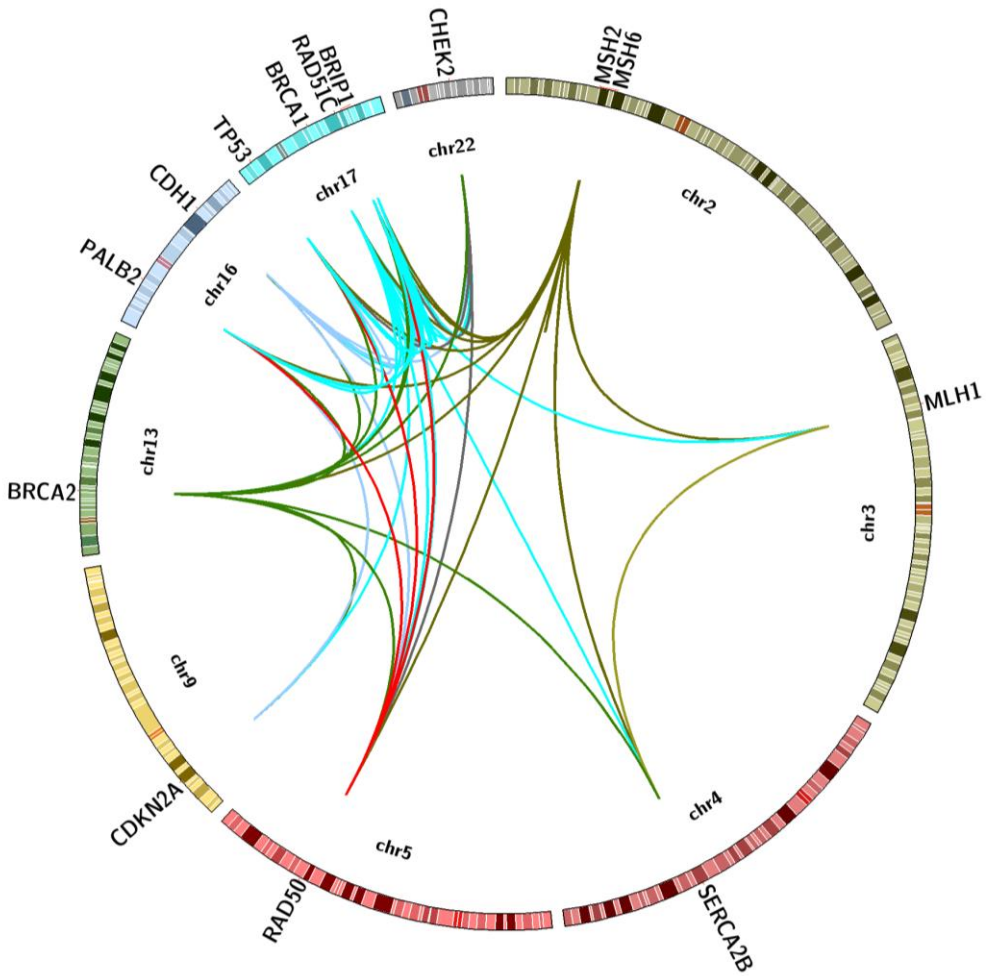


Figure 3.18: Small deletion distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with 14 genes attached to their corresponding chromosomes. The inner interconnections represent associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the corresponding chromosome.

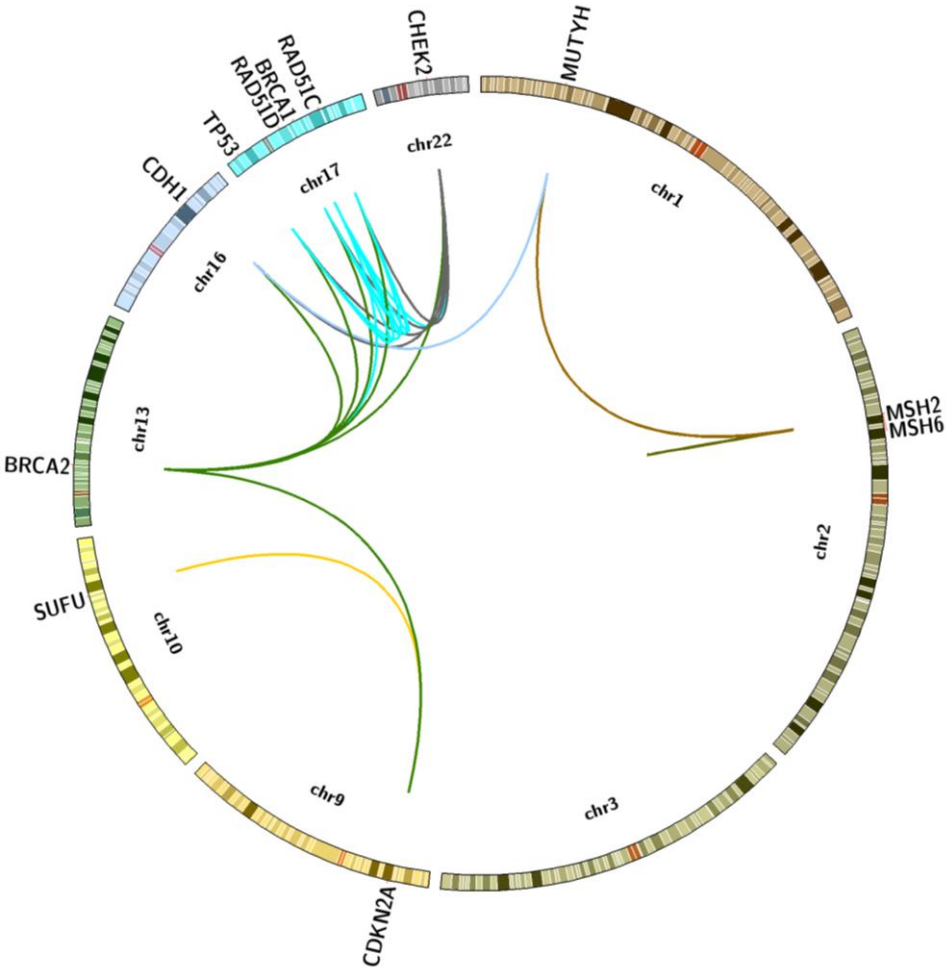


Figure 3.19: Splicing distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with the 12 genes attached to their corresponding chromosome the inner interconnections represent associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the corresponding chromosome.

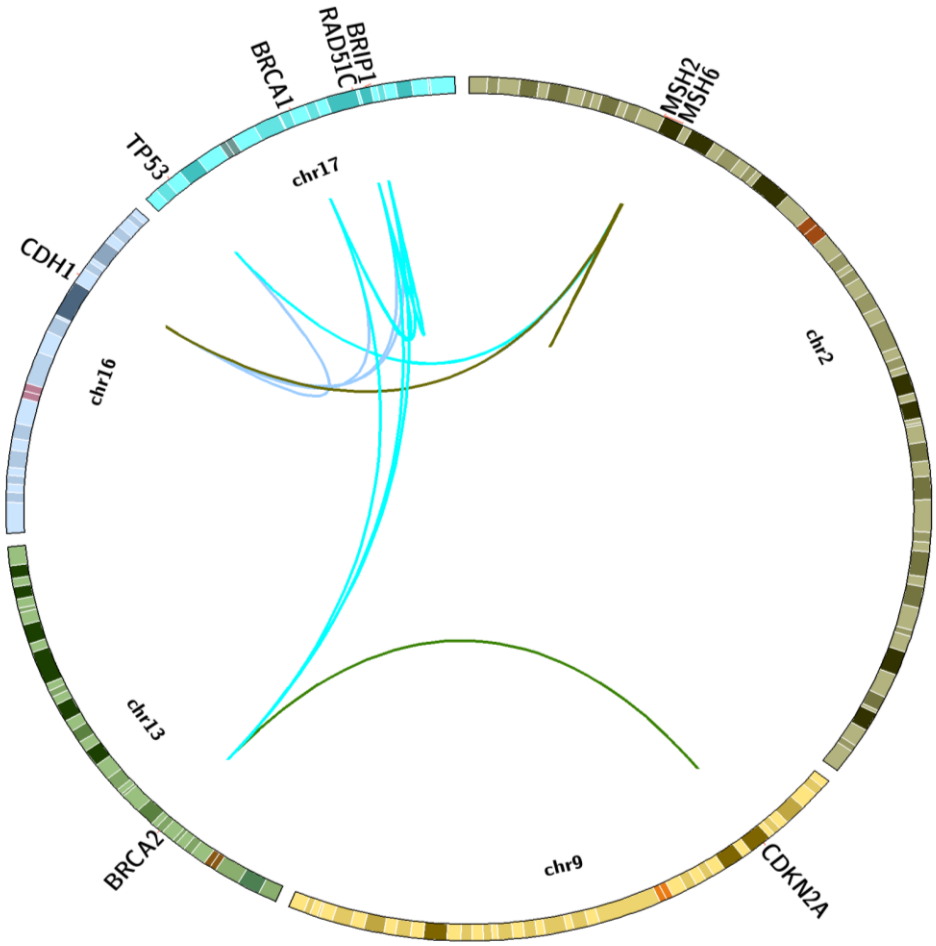


Figure 3.20: Small Insertions distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with 9 genes attached to their corresponding chromosome positions. The inner interconnections representing associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the colours of the corresponding chromosome.

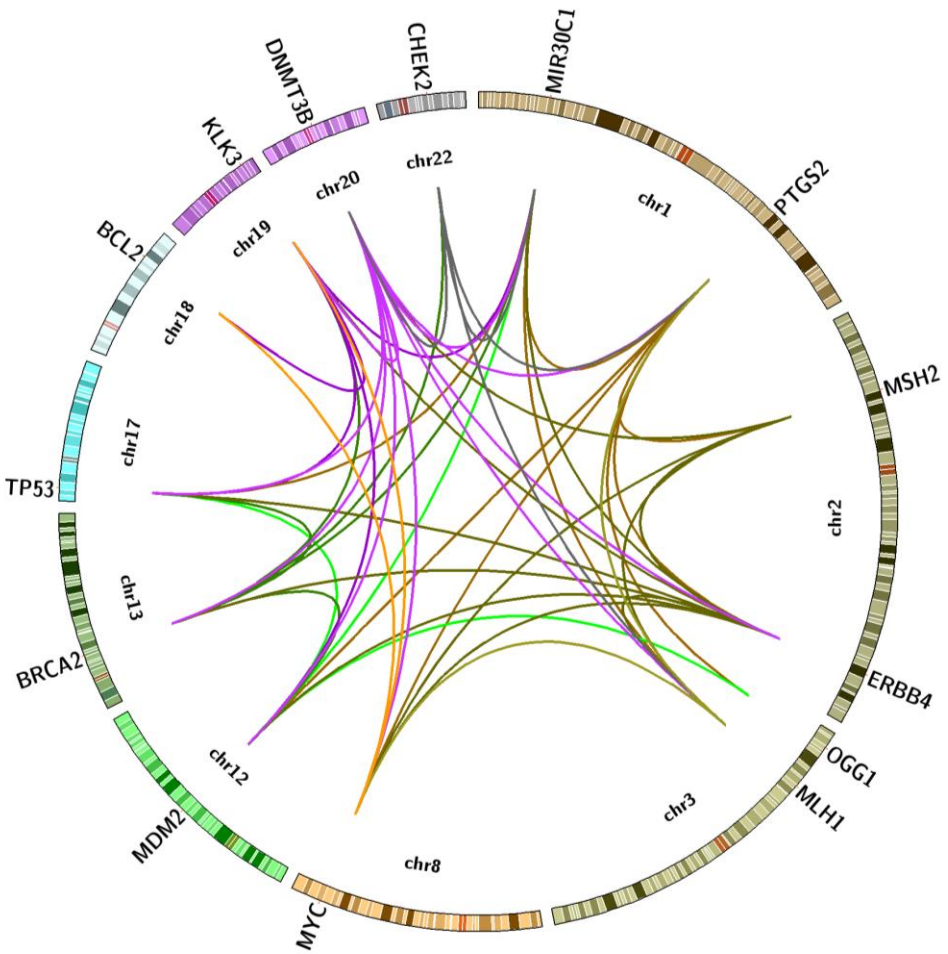


Figure 3.21: Regulatory mutation distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with 14 genes attached to their corresponding chromosomes. The inner interconnections represent associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the colours of the corresponding chromosome.

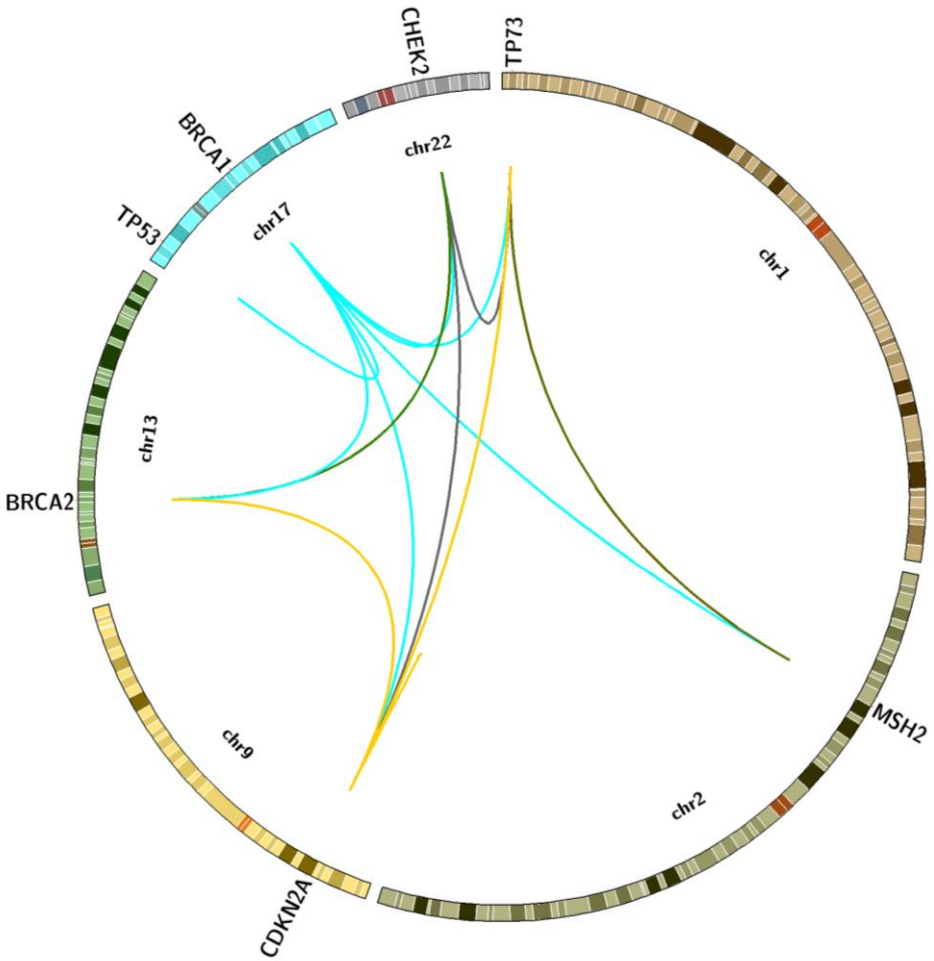


Figure 3.22: Gross Deletion distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with the 7 genes attached to their corresponding chromosomes. The inner interconnections represent associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the colours of the chromosome.

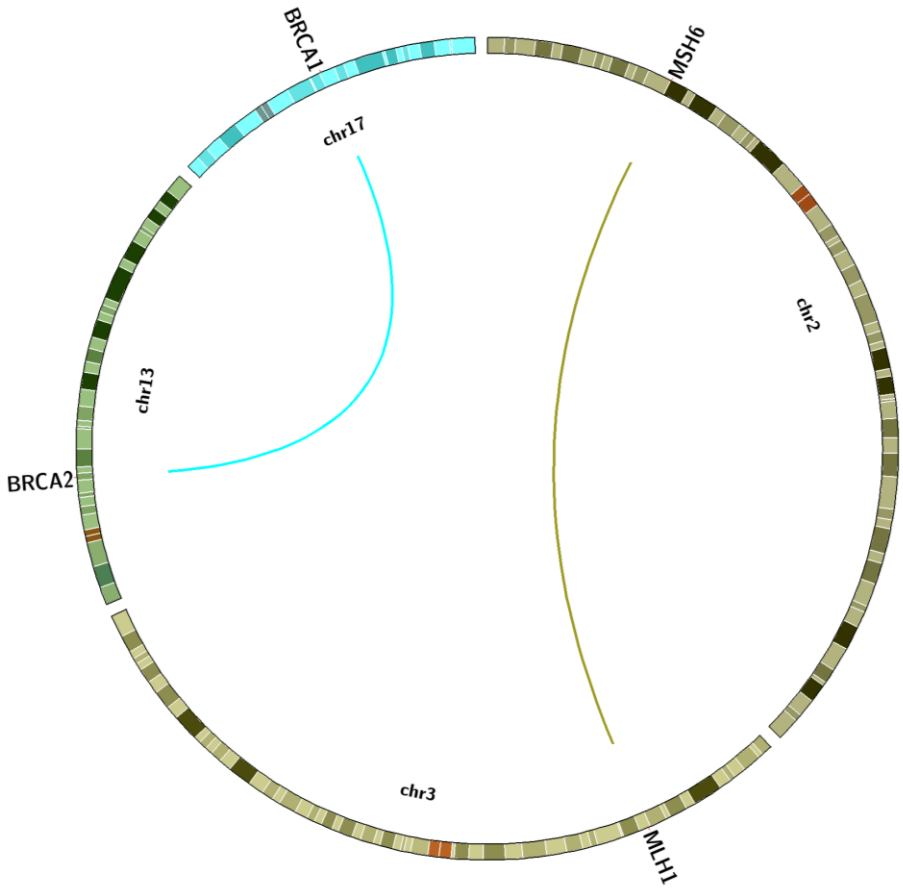


Figure 3.23: Small Indels distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with 4 genes attached to their corresponding chromosomes. The inner interconnections represent associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the colours of the corresponding chromosome.

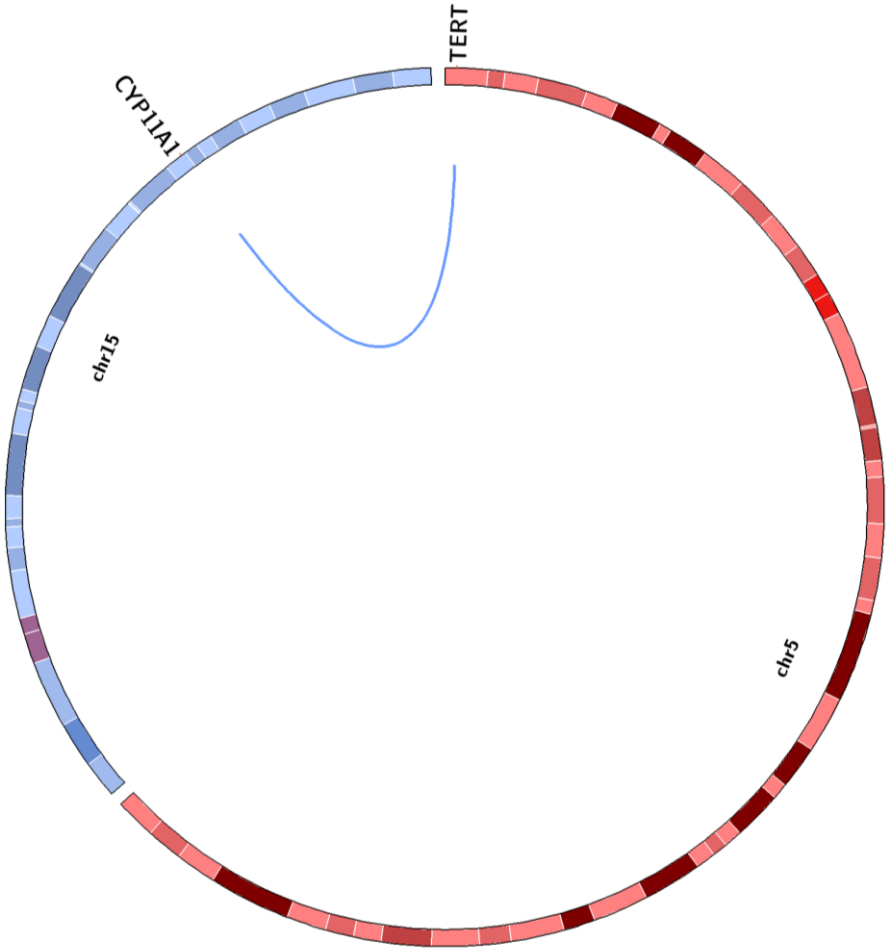


Figure 3.24: Repeated Variation distribution of Genome-wide mutations, where the blocks in the outer circle reflect the chromosomes, with the 2 genes attached to their corresponding chromosome. The inner interconnections representing associations between two or more genes connected if they contribute to same cancer. The colours of the interconnections represent the colours of the chromosome.

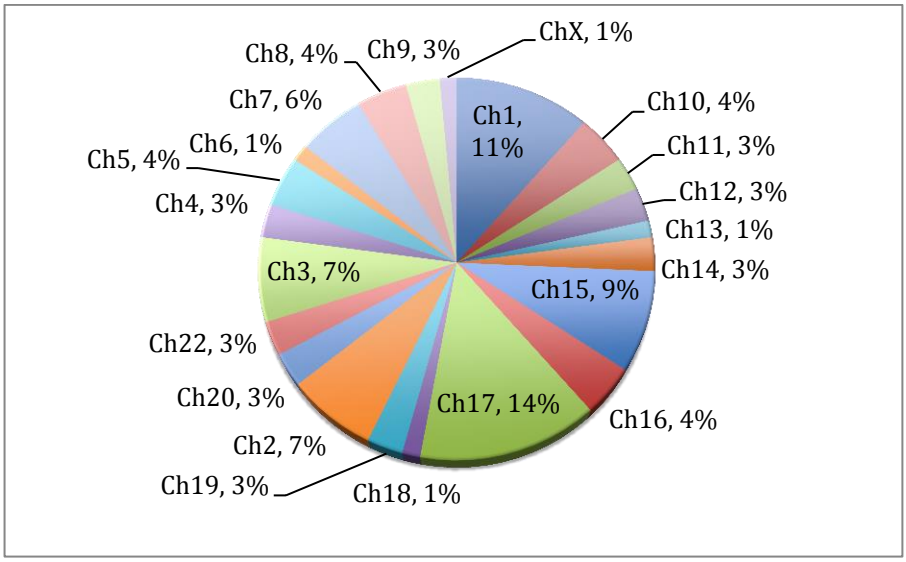


Figure 3.25: Distribution of the 69 involved genes in the Ggenome-wide distributions map for cancer genes over all human chromosomes.

3.2.3. KEGG pathways for the 69 linked Genes Identified in Genome-Wide Distribution sub-MAPs for each Mutation Class

The Kyoto Encyclopedia of Genes and Genome (KEGG) pathway database (Kanehisa et al., 2006) was used to identify the pathways represented in the Genome-wide Distributions sub-maps for each mutation class (Figures 3.17 to 3.24). This was carried by entering individual gene subsets of each Genome-wide Distribution sub-map (i.e. 10 mutation classes) into the pathway database.

The results showed that 66% of the missense nonsense and 71% of regulatory Genome-wide Distribution Map genes were mapped into pathways (Figure 3.26) and (Table 3.1), whereas other genome wide distribution genes showed no tendency to be in any pathways, and so the result was zero.

These findings indicate that missense nonsense and the regulatory mutation genes are involved in eight relevant pathways (1) Pathways in cancer, (2) Colorectal cancer, (3) Endometrial cancer, (4) Bladder cancer, (5) Mismatch repair, (6) P53 signalling pathways, (7) Cell cycle, and (8) Prostate cancer. Correspondingly, the missense nonsense and regulatory mutations showed for six identical cancers (1) Bladder cancer, (2) Bowel cancer, (3) Breast cancer, (4) Lung cancer, (5) Ovarian cancer, and (6) Prostate cancers in the constructed Genome-wide Distributions diagrams. This is an indication that these two classes of mutations could be the driver mutations in their interconnected genes to cause cancer (see table 3.1).

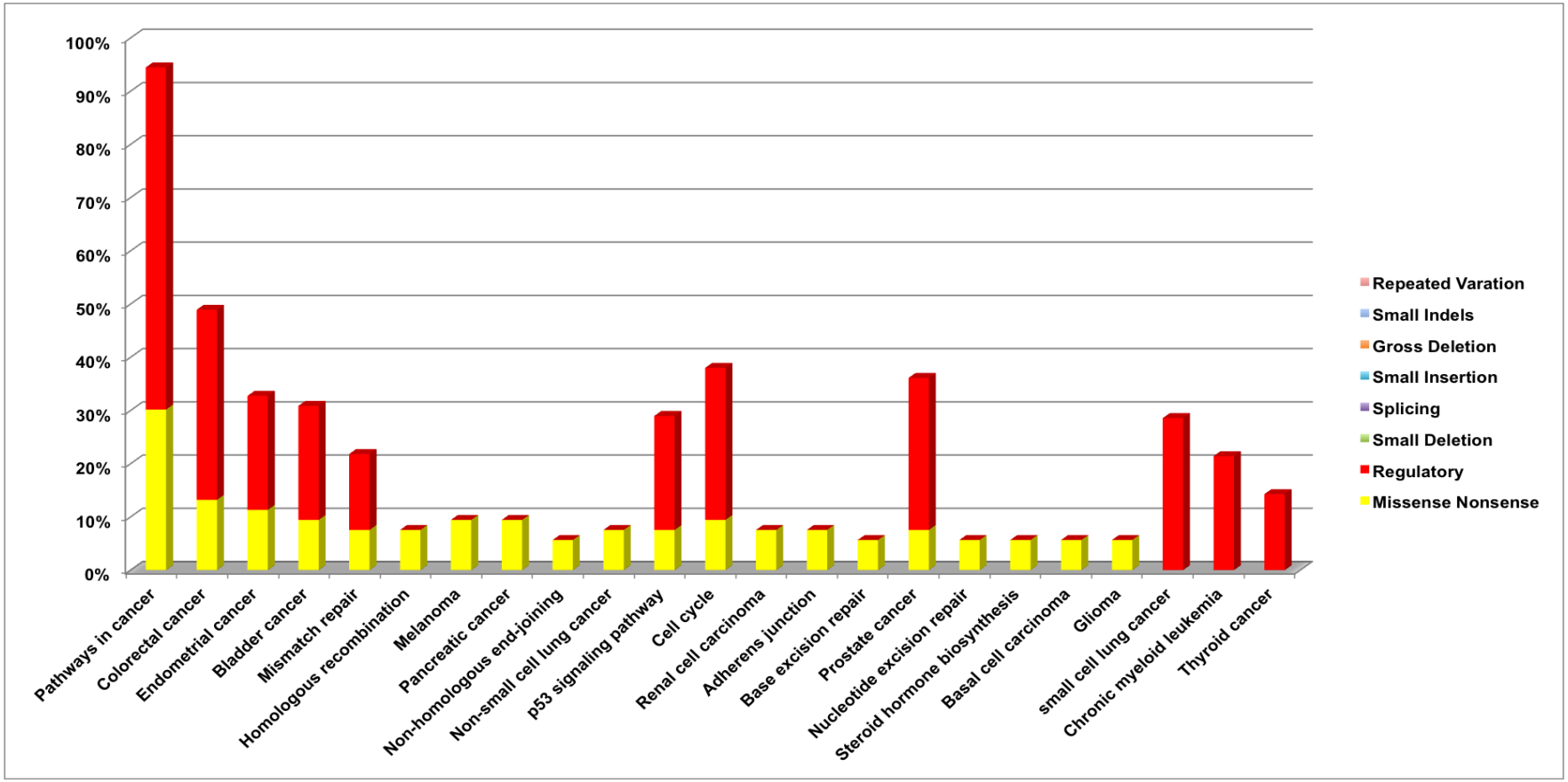


Figure 3.26: Pathways (KEGG) for Missense/Nonsense and regulatory mutations. The bar represents the percentage of genes involved in each pathway. The red is regulatory mutations and the yellow is a Missense/Nonsense mutation gene.

Table 3.1: Identified pathways for Missense/Nonsense and Regulatory mutation classes. For the involved genes in the construction of the Genome-wide Distribution Maps, The 1st column represents the identified pathway names (from KEGG). The 2nd column is the number of genes involved on the particular pathway for Missense/Nonsense mutation. The 3rd column is the number of genes involved in the particular pathway for regulatory mutation. The 4th column shows the gene symbols for the involved genes in the particular pathway for missense mutation. The 5th column shows the gene symbols for the involved genes in the particular pathway for regulatory mutation. The '0' means no genes were involved in that pathway, while 'NA' mean no available genes are reported for this KEGG pathway.

Pathway Term	Number of genes involved in a pathway (Missense/Nonsense)	Number of genes involved in a pathway (Regulatory)	Missense/Nonsense involved genes	Regulatory involved genes
Pathways in cancer	16	9	BMP4, EGFR, MSH6, AR, PTGS2, MSH2, VHL, ERBB2, MET, TP53, PML, MLH1, BRCA2, CDH1, CDKN2A, AXIN2	PTGS2, KLK3, MSH2, BCL2, TP53, MLH1, MDM2, BRCA2, MYC
Coloursectal cancer	7	5	EGFR, MSH6, MSH2, MET, TP53, MLH1, AXIN2	MSH2, BCL2, TP53, MLH1, MYC
Endometrial cancer	6	3	EGFR, ERBB2, TP53, MLH1, CDH1, AXIN2	TP53, MLH1, MYC
Bladder cancer	5	3	EGFR, CDKN2A, ERBB2, TP53, CDH1	TP53, MDM2, MYC
Mismatch repair	4	2	MSH6, MSH2, MLH1, MLH3	MSH2, MLH1
Homologous recombination	4	0	RAD51C, MRE11A, BRCA2, RAD50	NA
Melanoma	5	0	EGFR, CDKN2A, MET, TP53, CDH1	NA
Pancreatic cancer	5	0	EGFR, CDKN2A, ERBB2, TP53, BRCA2	NA
Non-homologous end-joining	3	0	XRCC4, MRE11A, RAD50	NA
Non-small cell lung cancer	4	0	EGFR, CDKN2A, ERBB2, TP53	NA
p53 signalling pathway	4	3	CDKN2A, TP53, CHEK2, ATM	TP53, MDM2, CHEK2
Cell cycle	5	4	CDKN2A, TP53, BUB1B, CHEK2, ATM	TP53, MDM2, CHEK2, MYC
Renal cell carcinoma	4	0	VHL, MET, FLCN, PTPN11	NA
Adherens junction	4	0	EGFR, ERBB2, MET, CDH1	NA
Base excision repair	3	0	XRCC1, PARP1, OGG1	NA
Prostate cancer	4	4	EGFR, AR, ERBB2, TP53	KLK3, BCL2, TP53, MDM2
Nucleotide excision repair	3	0	ERCC6, XPC, ERCC2	NA
Steroid hormone biosynthesis	3	0	CYP1B1, CYP1A1, CYP19A1	NA
Basal cell carcinoma	3	0	BMP4, TP53, AXIN2	NA
Glioma	3	0	EGFR, CDKN2A, TP53	NA
Small cell lung cancer	0	4	NA	PTGS2, BCL2, TP53, MYC
Chronic myeloid leukaemia	0	3	NA	TP53, MDM2, MYC
Thyroid cancer	0	0	NA	TP53, MYC

3.2.4. Functional Analytics of the Clustered Genes Groups

The 69 linked genes in the Genome-wide Distribution Map were clustered into three gene groups, based on the similarities of their shared biological functions (within DAVID; (Da Wei Huang and Lempicki, 2008). This generated gene-to-gene annotations, and outputted the result into a group of functionally-related genes. Using the kappa statistics method, only kappa agreement ≥0.35 is considered to reach the required threshold for biological function similarity.

Results from these analyses indicated:

(1) Clustered genes in Group 1 (i.e., *RAD50, MRE11A, NBN, MLH3, MSH6* and *MLH1*) show high interconnectivity as they are a) functionally-related genes (by DAVID), b) are shared by four cancers (i.e., Bowel, Breast, Ovarian and Womb by my Genome-wide Distribution Map), c) appear in the sub-maps of each of three mutation classes (i.e., missense/nonsense, small deletion and small indel mutations). In this way these genes are not just functionally-related but may also be cancer-related (see figure 3.27).

(2) Clustered Group 2 genes (i.e., *ERBB2, EGFR, MET* and *ERBB4*) while a) functionally-related (Via DAVID), b) less interconnected in terms of cancer-relatedness within my Genome-wide Distribution sub-maps, and c) only 2 genes (i.e. *EGFR* and *ERBB2*) were found to be cancer-related by sharing the same Missense/Nonsense mutation by Breast and Lung cancer (see figure 3.28),

(3) The clustered Group 3 genes (i.e., *CYP11A1, CYP19A1, CYP1A1, CYP1B1*) are a) functionally-related (via DAVID), b) three of these 4 genes (i.e. *CYP19A1, CYP1A1, CYP1B1*) show good interconnectivity in my Genome-wide Distribution sub-map, and c) were found to be cancer-related by sharing the same Missense/Nonsense mutation by Lung and Prostate cancers (see figure 3.29).

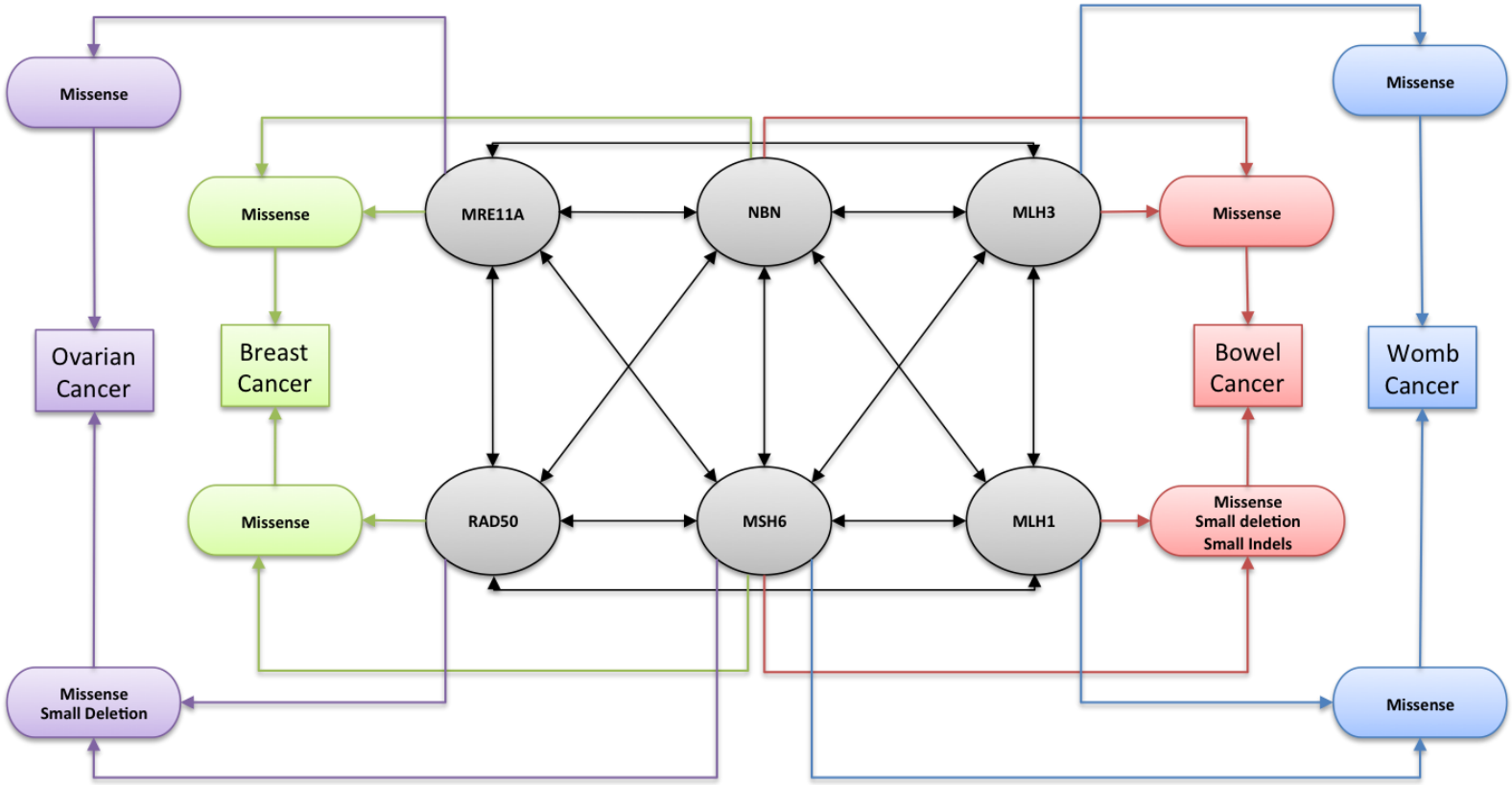


Figure 3.27: Functional clustering Group 1: Grey circles represent genes; black arrows connecting these genes on the bases of similarity threshold (i.e.≥ 0.35), for shared biological function; coloured arrows leaving each gene (black circles) and entering an oval (mutations) show the mutation class affecting that particular gene; arrows leaving an oval to enter a square show the cancer type associated with this specific mutation class.

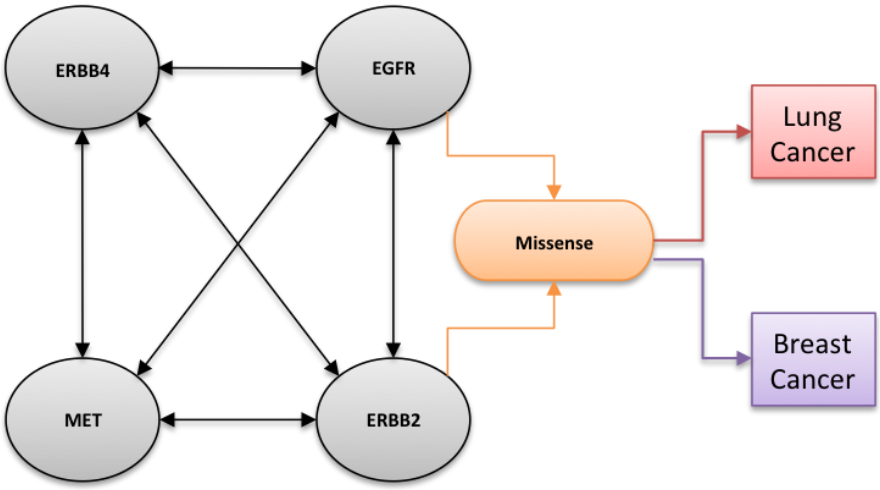


Figure 3.28: Functional clustering Group 2: Grey circles represent genes; black arrows connecting these genes on the bases of similarity threshold (i.e.≥ 0.35), for shared biological function; coloured arrows leaving each gene (black circles) and entering an oval (mutations) show the mutation class affecting that particular gene; arrows leaving an oval to enter a square show the cancer type associated with this specific mutation class.

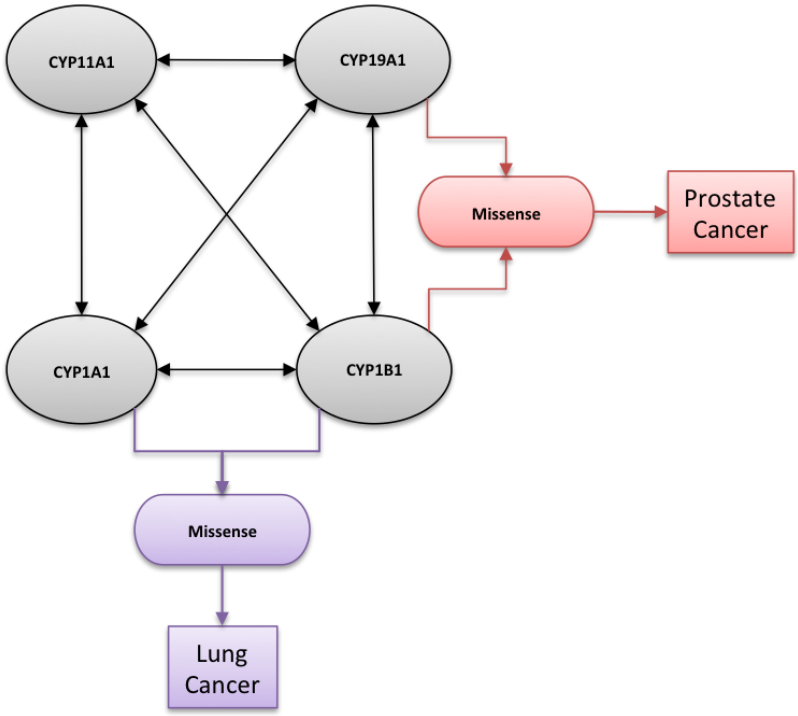


Figure 3.29: Functional clustering Group 3: Grey circles represent genes; black arrows connecting these genes on the bases of similarity threshold (i.e.≥ 0.35), for shared biological function; coloured arrows leaving each gene (black circles) and entering an oval (mutations) show the mutation class affecting that particular gene; arrows leaving an oval to enter a square show the cancer type associated with this specific mutation class.

3.3. Discussions/Conclusions

This chapter performed

(i) investigation into the influence of gene mutation classes, as a major biological factor in human cancers,

(ii) development of a human cancer map (HCM) based on genetic mutation classes data, and investigate the level of agreement (kappa) between interconnected cancer nodes

(iii) construction of a Genome-wide Distribution Map for cancer genes to identify KEGG pathways and thereafter to cluster genes into functionally-related groups based on their biological function.

Based on HCM, I explored the interconnectivity between cancer nodes in HCM sub-maps and noted some interesting features:

1. Bowel, Lung, Breast cancers presented as 3 central hubs that *all* connected to many other cancers (incl. melanoma, ovarian, stomach, prostate and brain cancers). Therefore, these three primary cancers represent major connecting cancer types in the HCM and may reflect the dominant position in these disorders in the human cancers. This is reflected by their position in the top 3 of most common cancers with a combined total lifetime risk of 1:15 as stated by Cancer Research UK (http://www.cancerresearchuk.org/cancer-info/cancerstats/incidence/risk/statistics-on-the-risk-of-developing-cancer).

2. The Missense/Nonsense class of mutations dominated the HCM, comprising a full 44% of all HCM interactions. By contrast, the remaining 9 mutation classes accounted for only 1%-14% of the HCM. However, while a Missense/Nonsense mutation appears to represent the most common cancer-inducing form of mutations, the kappa value for their interconnectivity was generally low.
3. By contrast, 5 other mutation classes (i.e., small deletions, splicing, small insertions, gross deletions and small indels) exhibited high kappa values (i.e., level of agreement) between connected nodes. This suggests that mutations of these types could be an important focus for future investigation into both the nature of cancer etiologies and associated potential therapeutic interventions.
4. The highest kappa values were found for 8 genes (i.e*., BRCA1, BRCA2, BRIP1, RAD51C, RAD51D, TP53, MLH1,* and *CDKN2A*), suggesting that these genes are particularly important connecting genes for 10 of the 29 (i.e., 34%) primary cancer nodes in HCM.

Based on Genome Wide Distribution Map,I explored the nature of the interconnected 69 cancer genes sub-maps (derived from the full 424 gene set), and found some interesting features:

1. Chromosome 17 carried the highest number of cancer-related genes (14% of total), with chromosome X carrying the fewest (1%).
2. Chromosomes 2,3,7,10,17 and 22 exhibited cancer-related gene mutations all the 10 classes.

3. The Missense/Nonsense Mutation type was the most common of all mutation classes, and was detected in all 22 chromosomes. Moreover, a mutation of the Regulatory Mutation type was detected in 11 of 22 chromosomes. Together these mutation types were most commonly associated with cancer-pathway genes.
4. By contrast, mutations of the Gross Insertion and Complex Rearrangement classes were not associated with cancer.
5. Six genes (i.e., *RAD50, MRE11A, NBN, MLH3, MSH6* and *MLH1*) exhibited similar biological functions, and were causative genes for Bowel, Breast, Ovarian and Womb cancer. All 4 cancers were associated with mutations of the same mutation classes.
6. Four genes (i.e., *ERBB4, MET, EGFR* and *ERBB2*) exhibited similar biological functions, and 2 of these (i.e., *EGFR* and *ERBB2*) were causative Breast and Lung cancers and were associated with mutations of the same mutation classes (i.e. Missense/Nonsense).
7. Four genes (i.e., *CYP11A1, CYP19A, CYP1A1* and *CYP1B1*) exhibited similar biological functions, and are causative genes for Lung and Prostate cancers.

In summary, the Missense/Nonsense mutation class may be a priority for researchers as this class is over-represented in the Human Cancer Map (HCM), Genome Wide Distribution Map, KEGG Pathway, Human genome, functional clustering of David Bioinformatics, and has a significant impact on human cancers.

Chapter Four

4. Integrated Data Analytics of a Protein-Protein Interaction Map with Missense/ Nonsense Mutations of corresponding Genes

Determination of the molecular cause of cancers has been a major focus of human genetic research since the early 1960s. The availability of advanced high-throughput genomic technologies and data has permitted large-scale genome-wide association studies (Cardon et al., 2007; Johnson and O'Donnell, 2009), and the analysis of complete DNA sequences of large numbers of cancer genomes to be conducted in order to gain understanding of how individual cancers develop (Stratton et al., 2009).

The aim of the work presented in this chapter was to associate genes based on the interactions of their encoded proteins. Specifically, gene sequences were interrogated for single base substitutions, which led to both a codon and amino acid change that was associated with cancer.

This approach was adopted to identify whether or not identical codon changes in two or more genes were associated with the same or different primary cancer type. For example, a codon change in the *ATM* gene, a switch from leucine to serine resulted in breast cancer, whereas the same codon change in the *MSH2* gene resulted in bowel cancer.

In order to gain greater knowledge of the associations between multiple genes related to cancer, and their single base substitutions I sought to:

(i) Construct a Protein-Protein Interaction Map (PPIM) derived from 424 cancer-associated gene set using the Biological General Repository for Interaction Datasets (BioGRID) (www.thebiogrid.org).

(ii) Extract Missense/Nonsense mutations (i.e. from codons with resultant amino acid change) for cancers.

(iii) Investigate the distribution of changes in all 21 amino acids in relation to cancers

4.1. Protein-Protein Interaction Map (PPIM)

A PPIM was constructed to investigate single base substitutions for the 424 gene set using BioGRID (Stark et al., 2011) and the GeneMANIA Cytoscape plugin (Montojo et al., 2010). Two genes were associated if they were found to interact with each other in a protein-protein interaction study. Here only the physical interactions of proteins were considered. These associating genes were then visualized using Gephi tools (Bastian et al., 2009) (see figure 4.1). The PPIM contains nodes that correspond to genes, the size of which (nodes) is proportional to the number of distinct genes connecting to it, while interconnected links (also called 'edges') represent the physical interaction of genes. The PPIM I obtained consisted of 292 genes with a total of 4,024 physical interactions among the corresponding proteins of genes. Certain genes tend to form a central hub connecting many other genes, where the distribution degree (k) is high. For example, TP53 (k=72), BRCA1 (k=55), CREBBP (k=56), ESR1 (k=58), AR (k=27), ATM (k=22), BRCA2 (k=13) and RAD41 (k=28) all connect many other genes. These data indicate that each of these genes is highly interacted with others genes that encode proteins in the PPIM. However, a focus here was to identify Missense/Nonsense mutations in the DNA for every gene (that encoded a PPIM protein), whereby a single nucleotide causes a codon change, with a resultant unexpected amino acid alteration, that was associated with cancer.

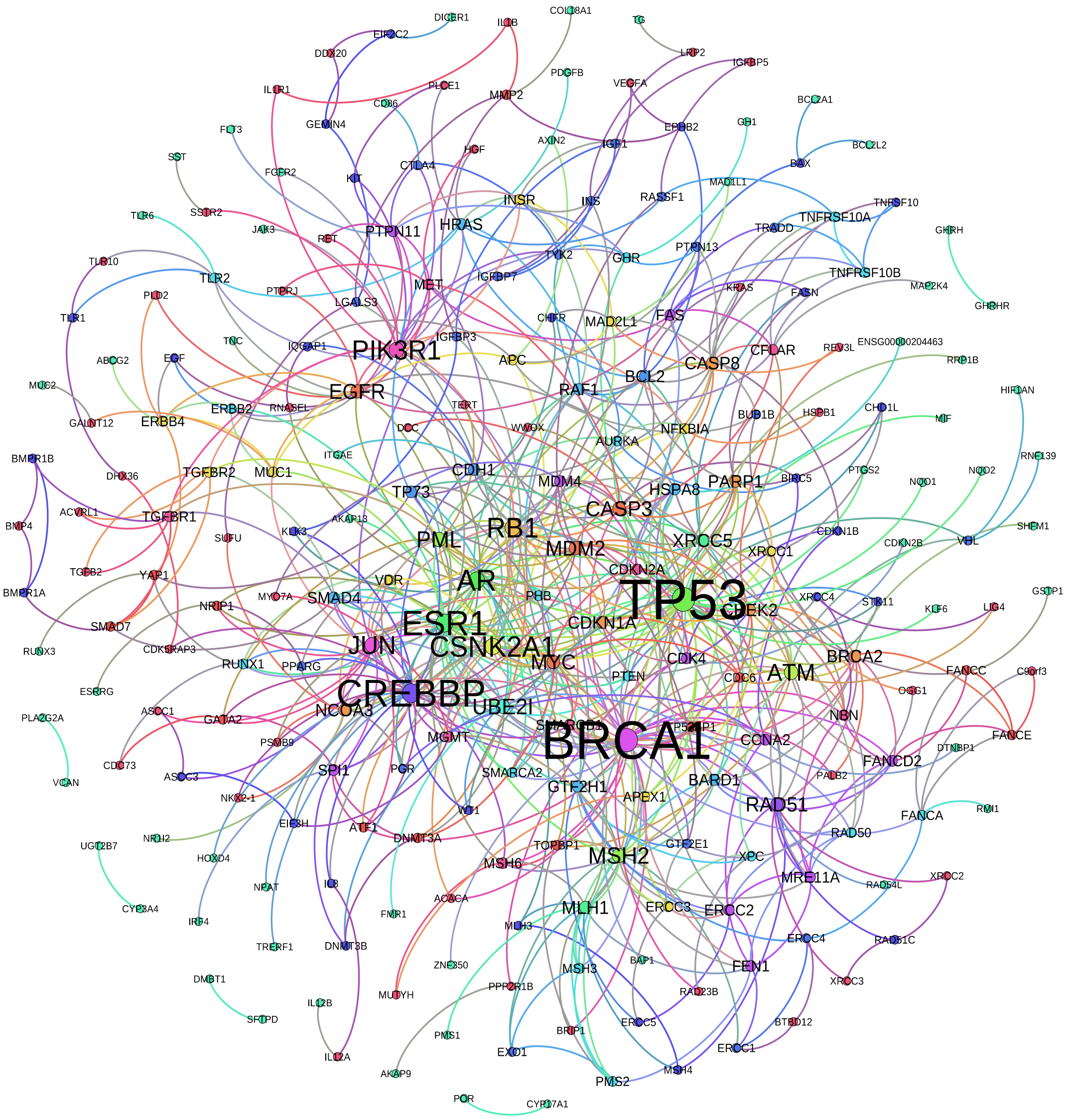


Figure 4.1: Protein–Protein Interaction Map (PPIM): Nodes represent unique proteins, and the links (edges) identify the protein-protein physical interactions. The colour of the link indicates the originating or source node in the physical protein-protein interaction.

4.1.1. Interrogation of Missense/Nonsense Mutations of Codons and their encoded Amino Acid Changes in a PPIM-Informed 292 Cancer-Associated Gene set

The extraction of Missense/Nonsense mutations involving single nucleotide changes, which resulted in cancer-associated amino acid changes, was investigated manually for the 292-gene set informed by the PPIM. This was conducted one base at a time within the mutation using complementary data sources e.g., PubMed, OMIM (Amberger et al., 2009), UniProtKB/Swiss-Port (Boutet et al., 2007) and HGMD(Stenson et al., 2009). The purpose of this Missense/Nonsense mutations data extraction was to detect:

(i) the single base change in the affected codon

(ii) the amino acid resulting from the altered codon

(iii) the peptide position of the altered amino acid

(iv) the associated cancer disorder(s) resulting from the altered amino acid using PubMed for every gene.

In total, 2,851 Missense/Nonsense mutation records were found for 161 of the 292 interrogated genes and were stored as independent records for each gene symbol (see appendix 3 table 3.1). Thus, 131 genes informed by the PPIM did not have associated Missense/Nonsense mutation records. Thus, I was interested in next assessing whether identical amino acid alterations (e.g., a Ser to Leu, Ala to Leu, Tyr to Leu…. etc switch) for two or more genes within the 161-gene set were associated with the same primary cancer type.

4.1.2. Do Identical Amino Acid Alterations for two or more Genes Associated with the same Primary Cancer Type?

To address whether identical amino acid alterations for two or more genes within the 161 genes set associate with the same primary cancer type, I combined the Missense/Nonsense mutation records with the map of interconnected gene nodes that informed the PPIM. This was achieved by:

(i) Dividing the extracted Missense/Nonsense mutation records (table 3.1 appendix 3) into sub-tables, each consisting of records for one altered amino acid that was related to a primary cancer type. In total 21 tables were constructed for each of the 21 amino acid by running a string using R language.

(ii) Associating the newly-constructed amino acid sub-table records with the PPIM connected nodes. This association was verified as follows. For two or more genes to be considered connected they must exhibit a physical interaction in the PPIM (recognized in BioGrid) and have an identically-altered amino acid that is also associated with the same primary cancer type. For example, the following two genes (*TP53* and *MSH2*) are said to be associated where a single nucleotide change in *TP53* changes the ‘GGC’ codon to ‘GAC’ resulting in a Glycine’ to ‘Aspartate’ amino acid

alteration to cause ‘Bowel cancer’, and a single nucleotide change in *MSH2*, resulting in a codon change from ‘TAT’ to ‘GAT’, resulting in a ‘Tyrosine’ to ‘Aspartate’ amino acid change also causes ‘Bowel cancer’.

(iii) Tagging the associated gene records by 0,1 or 2 where ‘0’ represents the same altered amino acid, ‘1’ the identical change in codon, and ‘2’ for the identical encoded amino acid change in the identical sequence position of the resultant peptide (appendix 3 table A.3.2).

Due to the complex nature of biological data a manual verification process was carried out for combining and verifying each record successively, using the Gephi visualizing tools. This generated 21 ‘altered amino acid’ network maps involving 57 genes associated with 8 primary cancers (figures 4.2 to 4.22).

As a navigation guide to the following 21 maps, the reader should note that:

a) The colour of gene nodes represents the number of links (edges i.e. physical interactions). For any connecting two genes a blue edge represents the same altered amino acid (regardless of the nature of the codon change); a red edge represents identical codon change; and a green edge represents identical peptide sequence position of the altered amino acid.

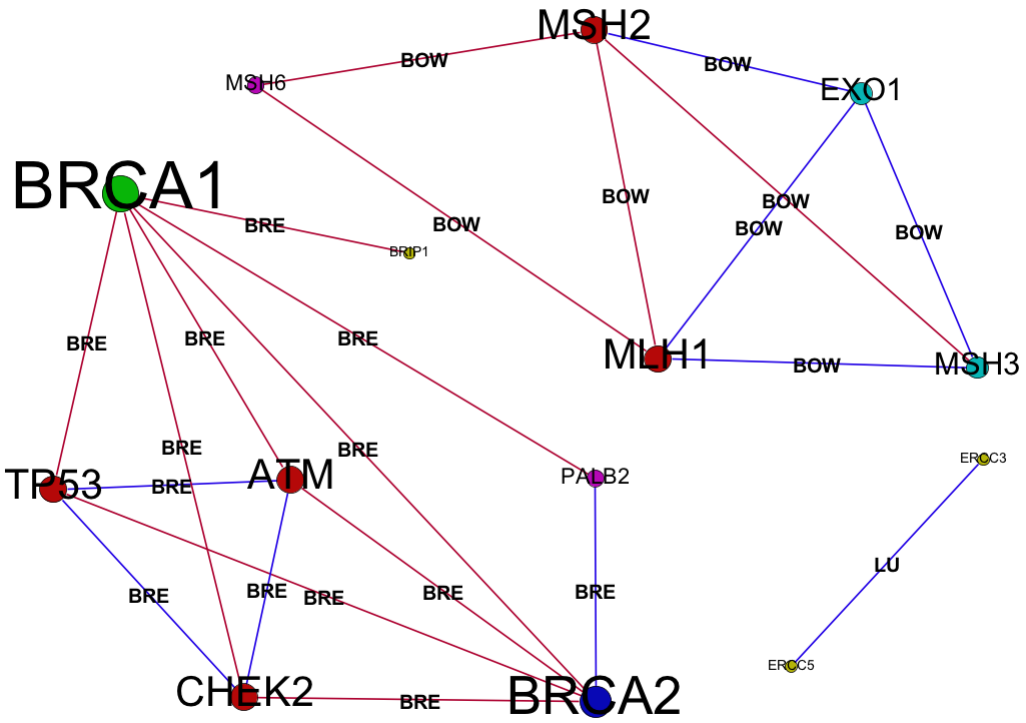


Figure 4.2: PPIM of an **Alanine** altered amino acid and associated with the same primary cancer. 59% of the edges (red) show identical codon changes, while 41% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 59% of the connected edges, Bowel cancer (BOW) involves 36% of the connected edges and Lung (LU) involves 6% of the connected edges).

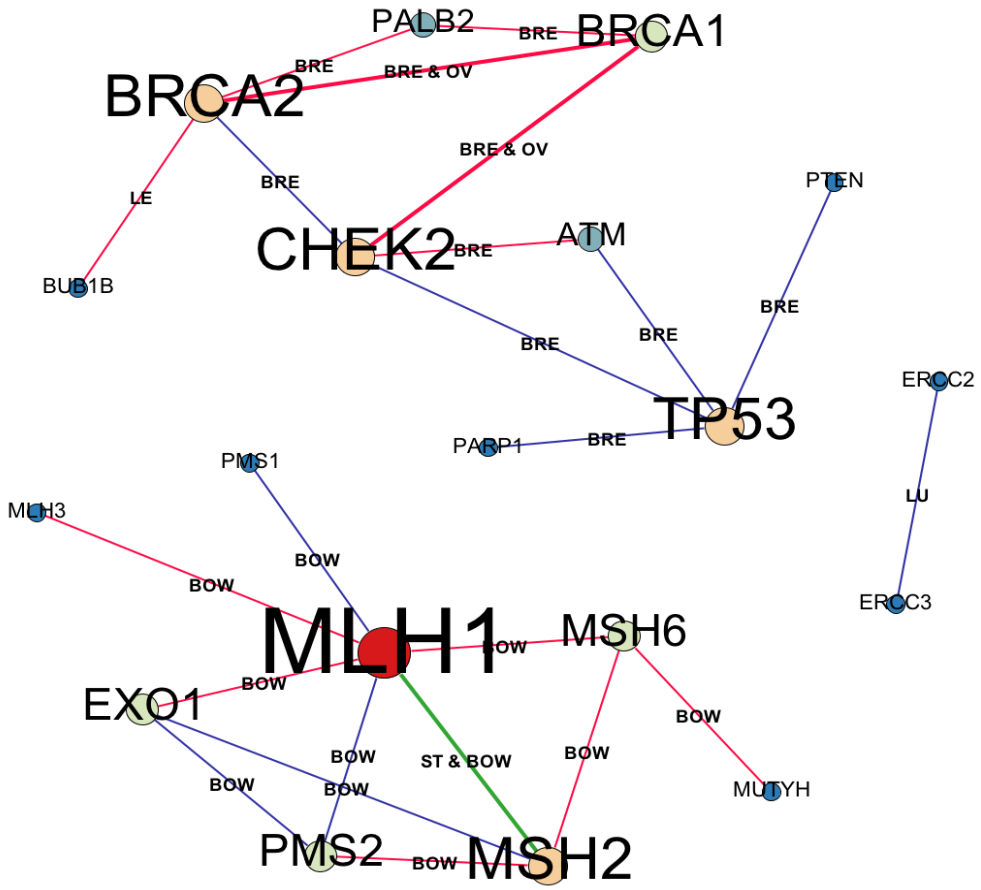


Figure 4.3: PPIM of an **Arginine** altered amino acid and associated with the same primary cancer. 52% of the edges (red) show identical codon changes, while 43% of the edges (blue) show non-identical codon changes and the remaining 5% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 38% of the connected edges, Bowel cancer (BOW) involves 43% of the connected edges, Lung cancer (LU) involves 4% of the connected edges, Ovarian cancer (OV) involves 9% of the connected edges and Stomach cancer (ST) involves 4% of the connected edges).

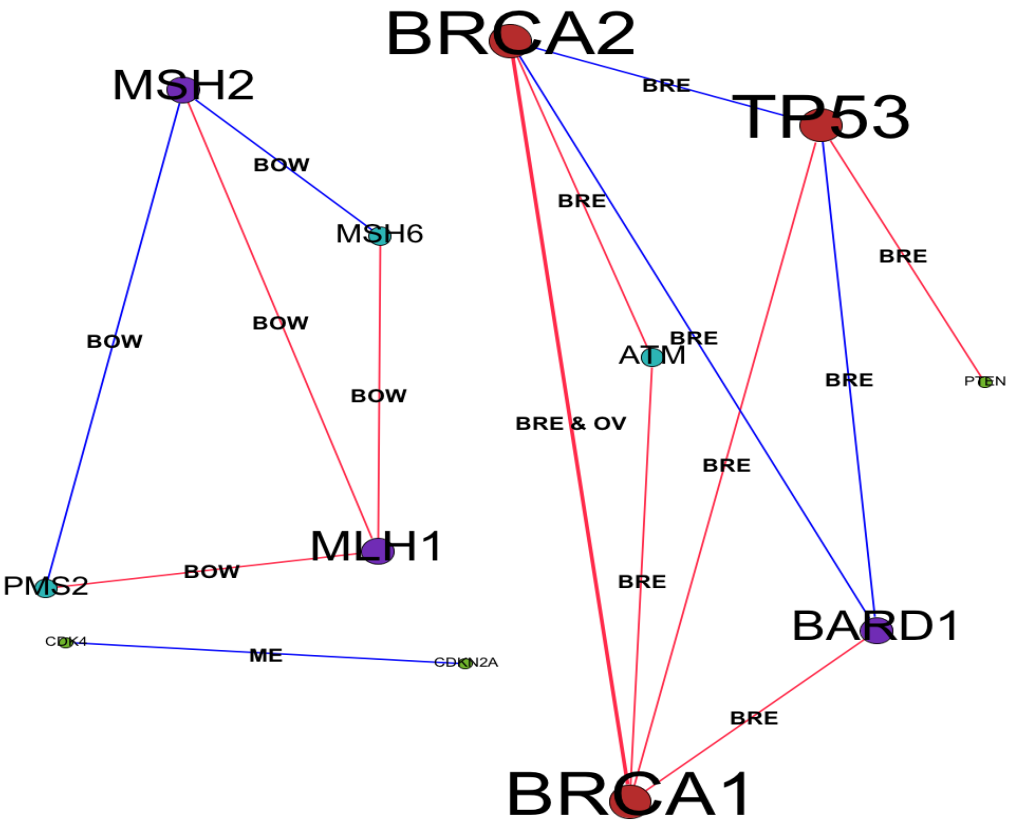


Figure 4.4: PPIM of an **Asparagine** altered amino acid and associated with the same primary cancer. 60% of the edges (red) show identical codon changes, while 40% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 53% of the connected edges, Bowel cancer (BOW) involves 33% of the connected edges, Ovarian cancer (OV) involves 7% of the connected edges and Melanoma (ME) involves 7% of the connected edges).

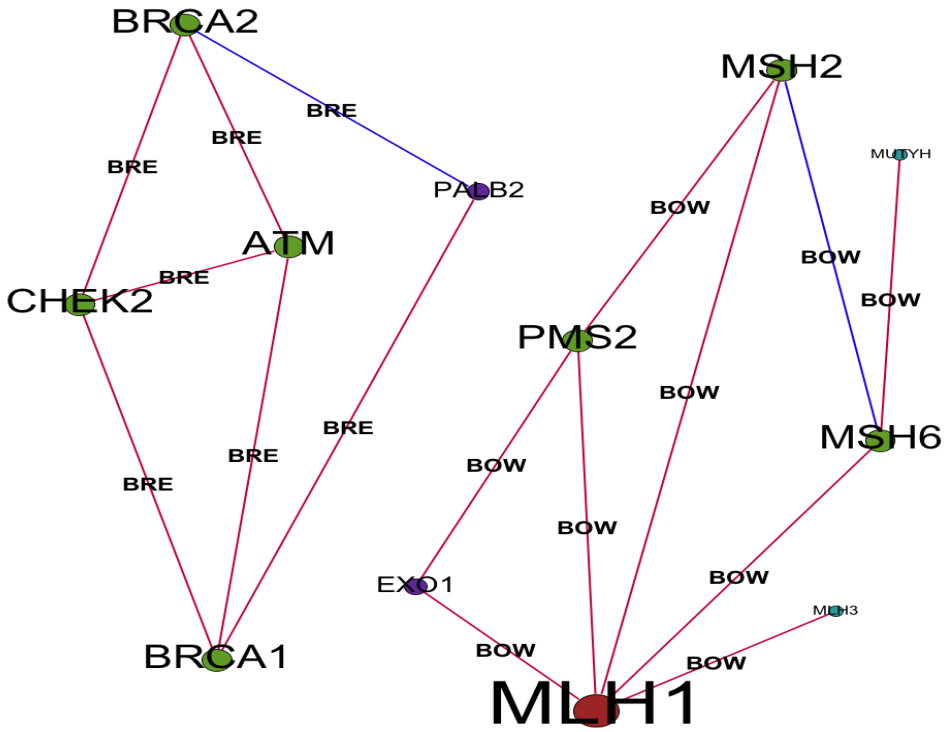


Figure 4.5: PPIM of a **Glutamate** altered amino acid and associated with the same primary cancer. 88% of the edges (red) show identical codon changes, while 12% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 44% of the connected edges and Bowel cancer (BOW) involves 56% of the connected edges).

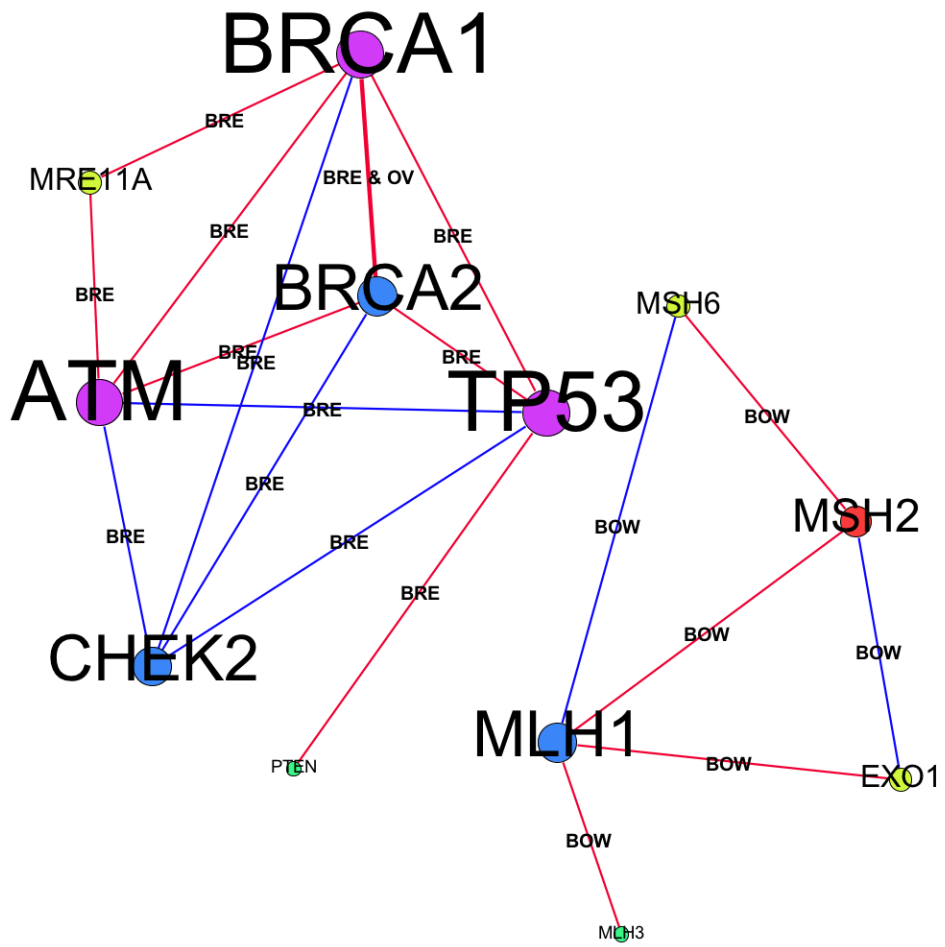


Figure 4.6: PPIM of a **Glycine** altered amino acid and associated with the same primary cancer. 63% of the edges (red) show identical codon changes, while 37% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 63% of the connected edges, Bowel cancer (BOW) involves 32% of the connected edges and Ovarian cancer (OV) involves 5% of the connected edges).

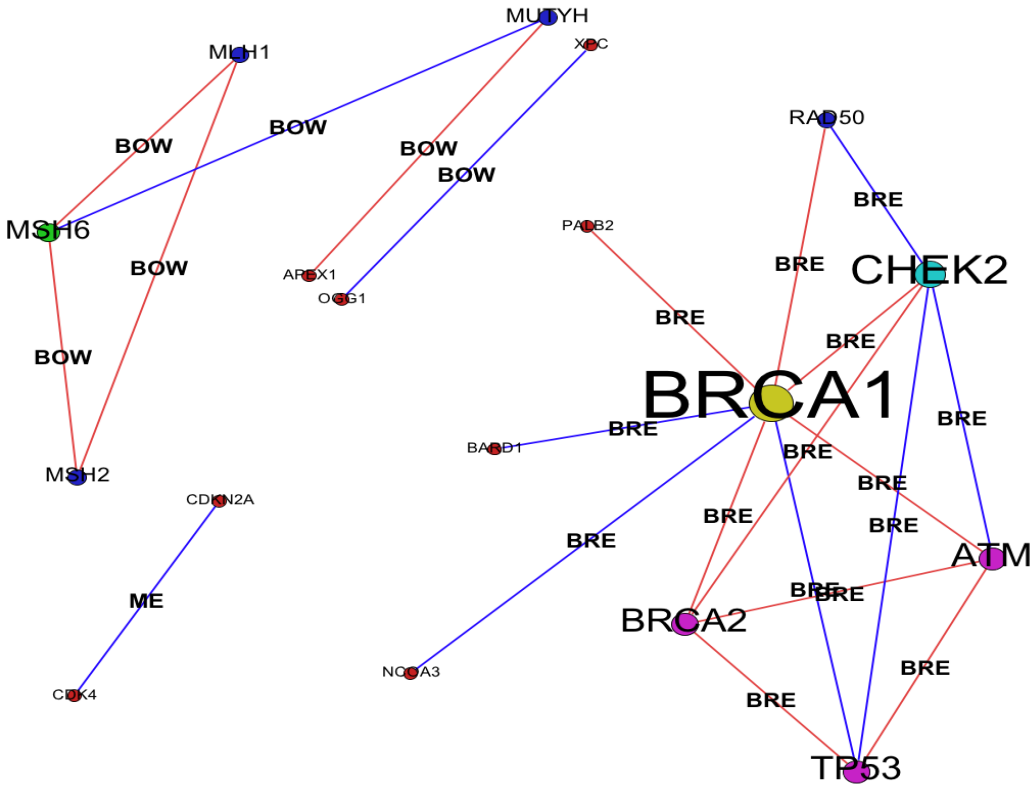


Figure 4.7: PPIM of a **Histidine** altered amino acid and associated with the same primary cancer. 55% of the edges (red) show identical codon changes, while 36% of the edges (blue) show non-identical codon changes and the remaining 9% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 68% of the connected edges, Bowel cancer (BOW) involves 27% of the connected edges and Melanoma (ME) involves 5% of the connected edges).

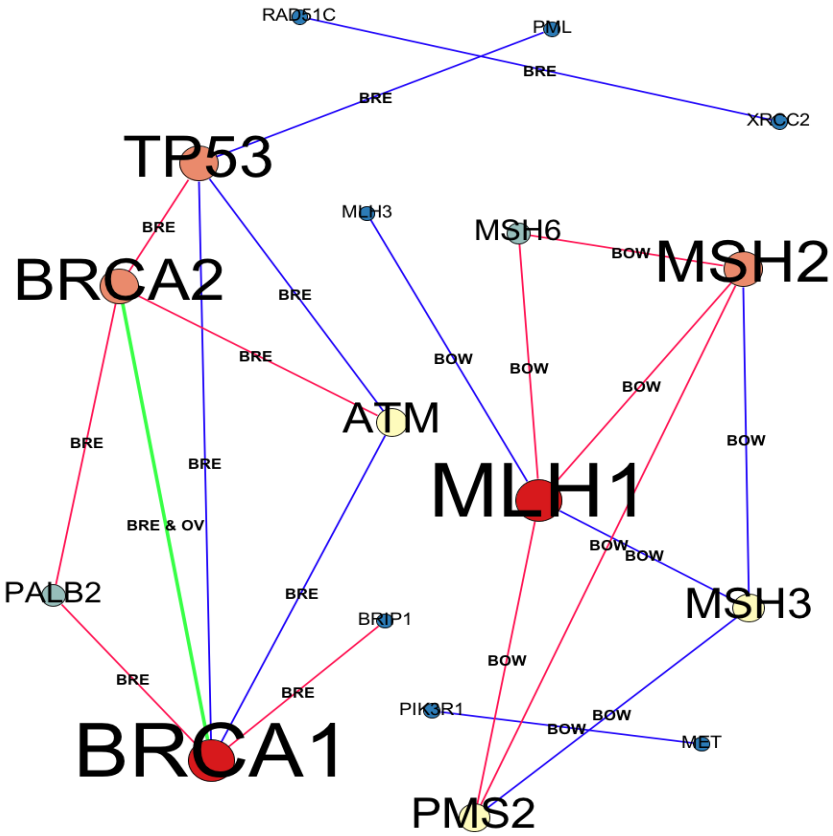


Figure 4.8: PPIM of an **Isoleucine** altered amino acid and associated with the same primary cancer. 48% of the edges (red) show identical codon changes, while 48% of the edges (blue) show non-identical codon changes and the remaining 4% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 48% of the connected edges, Bowel cancer (BOW) involves 48% of the connected edges and Ovarian cancer (OV) involves 4% of the connected edges).

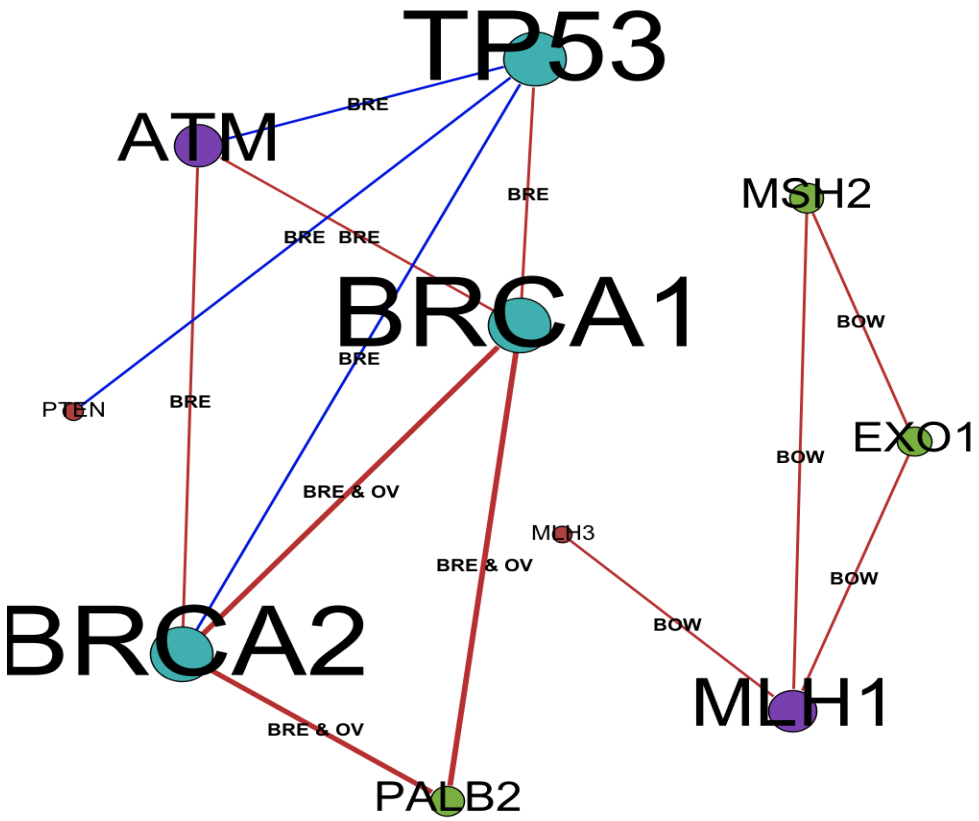


Figure 4.9: PPIM of a **Lysine** altered amino acid and associated with the same primary cancer. 77% of the edges (red) show identical codon changes, while 23% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 46% of the connected edges, Bowel cancer (BOW) involves 31% of the connected edges and Ovarian cancer (OV) involves 23% of the connected edges).

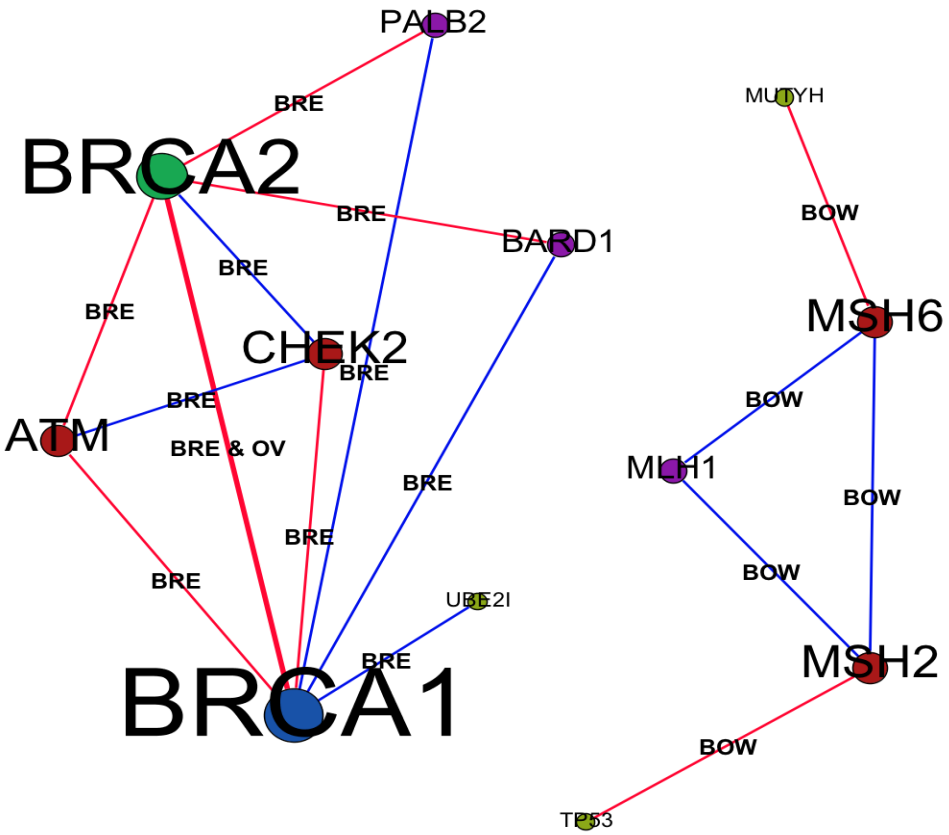


Figure 4.10: PPIM of a **Methionine** altered amino acid and associated with the same primary cancer. 50% of the edges (red) show identical codon changes, while 50% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 63% of the connected edges, Bowel cancer (BOW) involves 31% of the connected edges and Ovarian cancer (OV) involves 6% of the connected edges).

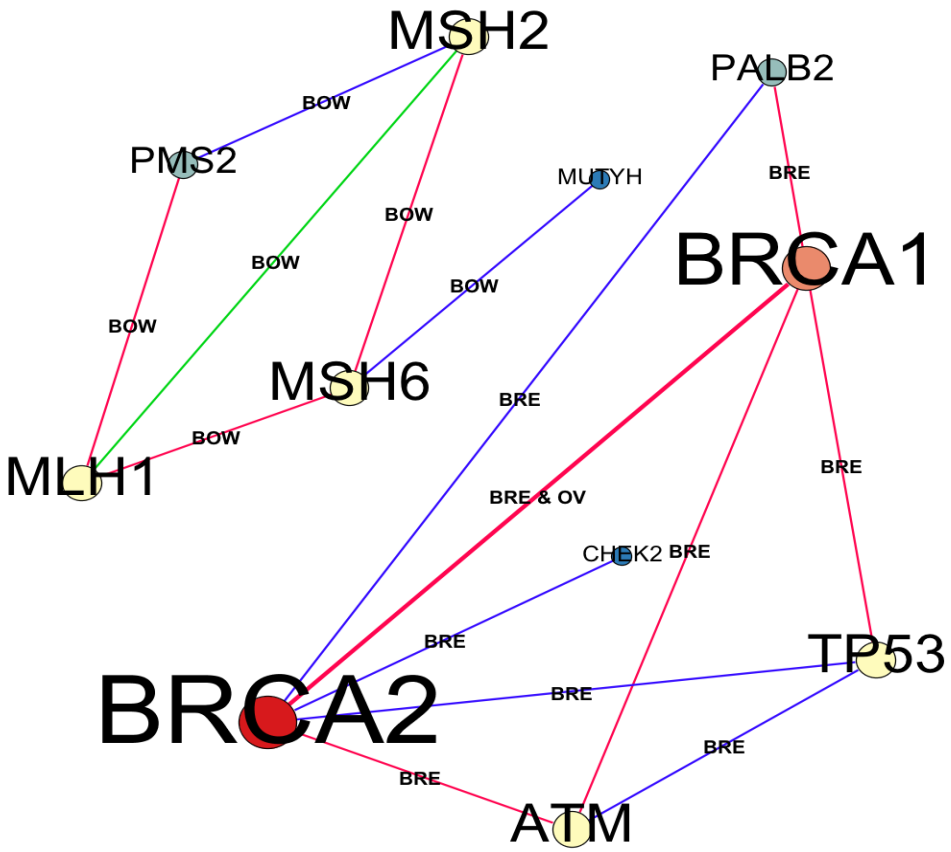


Figure 4.11: PPIM of a **Proline** altered amino acid and associated with the same primary cancer. 53% of the edges (red) show identical codon changes, while 40% of the edges (blue) show non-identical codon changes and the remaining 7% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 54% of the connected edges, Bowel cancer (BOW) involves 40% of the connected edges and Ovarian cancer (OV) involves 6% of the connected edges).

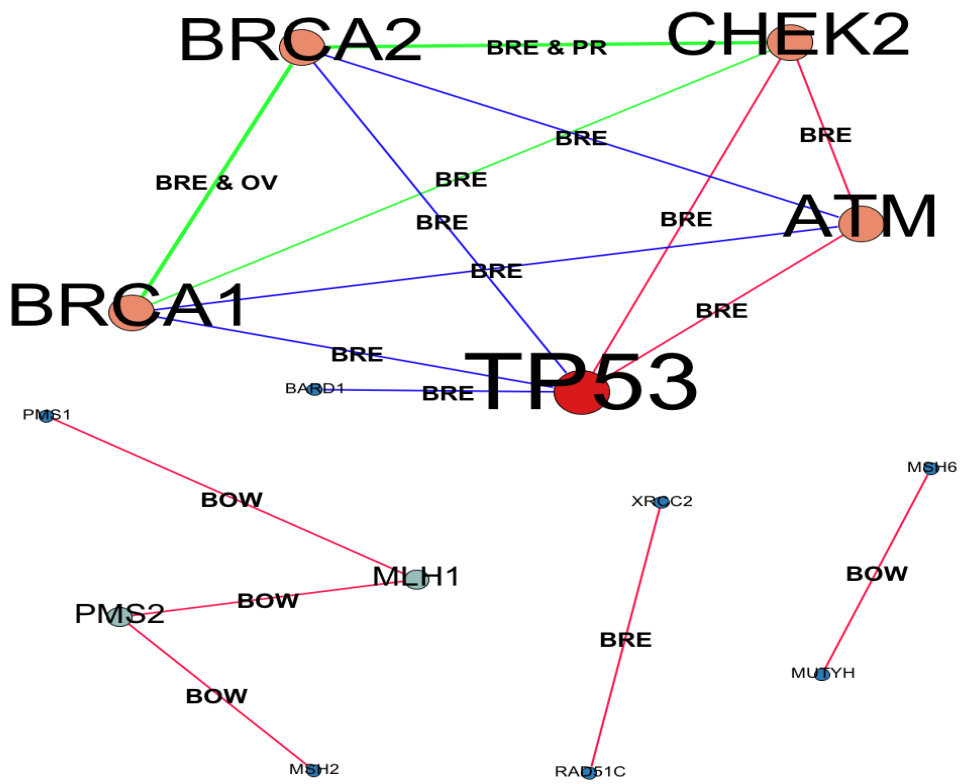


Figure 4.12: PPIM of a **Termination (stop codon)** altered amino acid and associated with the same primary cancer. 50% of the edges (red) show identical codon changes, while 31% of the edges (blue) show non-identical codon changes and the remaining 19% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 63% of the connected edges, Bowel cancer (BOW) involves 25% of the connected edges, Ovarian cancer (OV) involves 6% of the connected edges and Prostate cancer (PR) involves 6% of the connected edges).

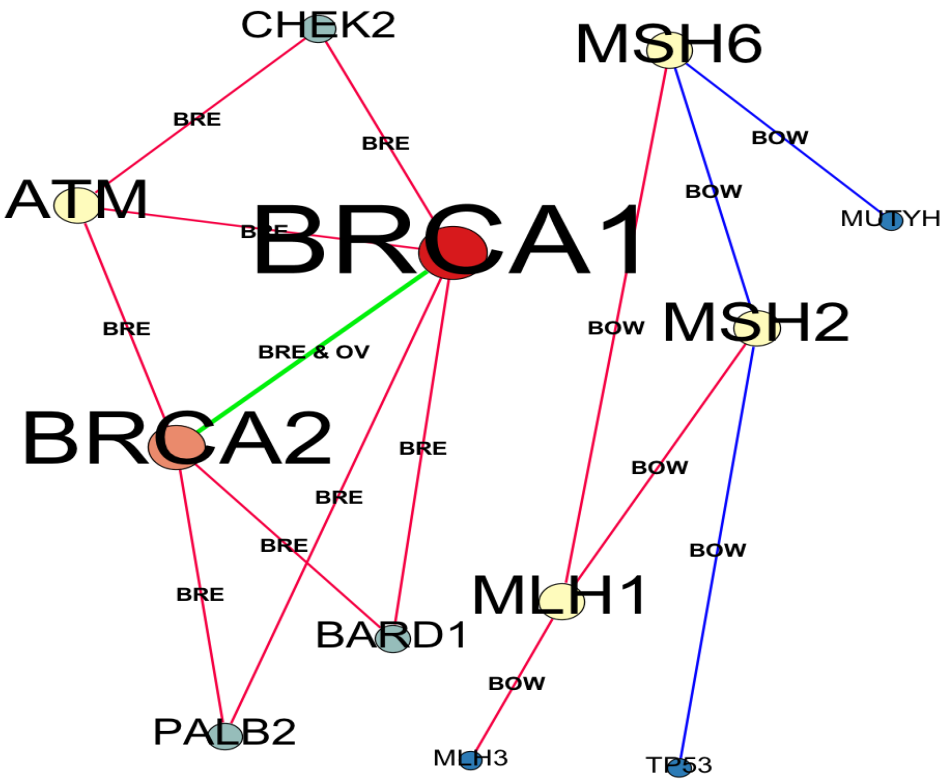


Figure 4.13: PPIM of a **Threonine** altered amino acid and associated with the same primary cancer. 73% of the edges (red) show identical codon changes, while 20% of the edges (blue) show non-identical codon changes and the remaining 7% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 53% of the connected edges, Bowel cancer (BOW) involves 40% of the connected edges and Ovarian cancer (OV) involves 7% of the connected edges).

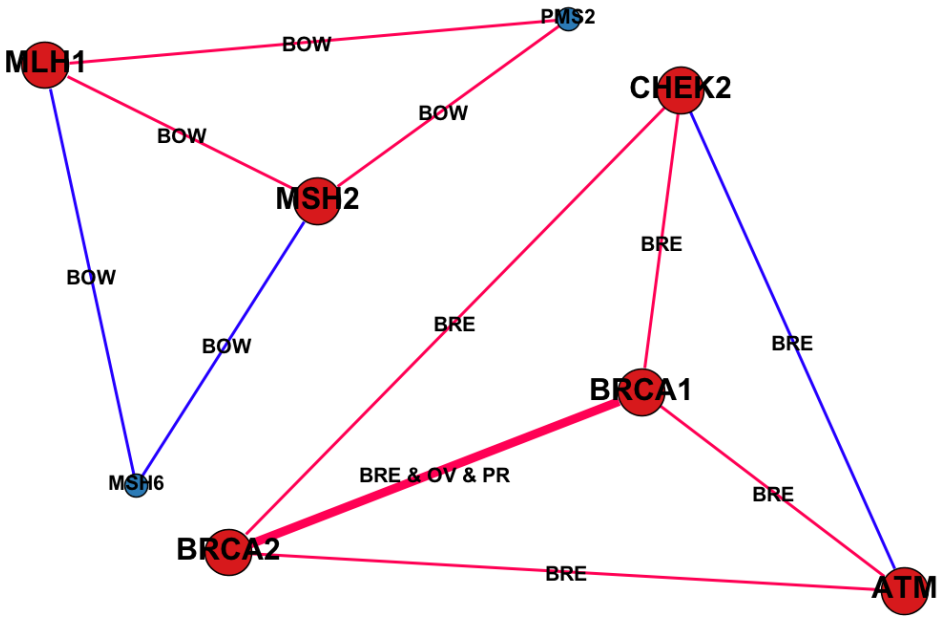


Figure 4.14: PPIM of a **Tyrosine** altered amino acid and associated with the same primary cancer. 73% of the edges (red) show identical codon changes, while 27% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 45% of the connected edges, Bowel cancer (BOW) involves 45% of the connected edges and Ovarian cancer (OV) involves 5% of the connected edges).

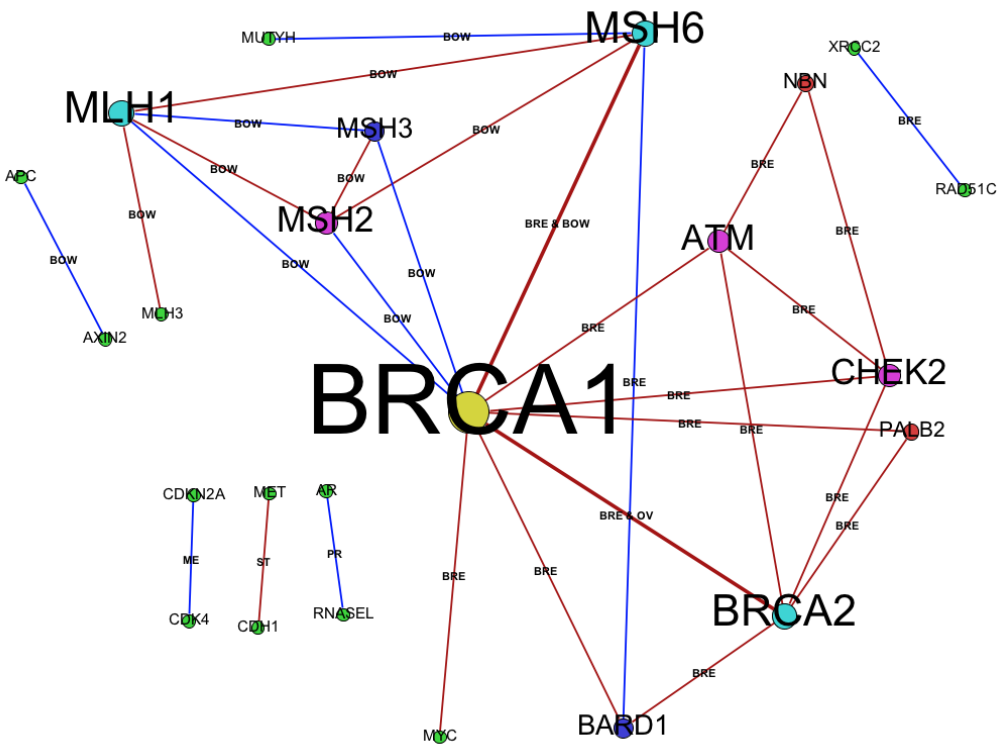


Figure 4.15: PPIM of a **Serine** altered amino acid and associated with the same primary cancer. 67% of the edges (red) show identical codon changes, while 33% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 47% of the connected edges, Bowel cancer (BOW) involves 37% of the connected edges, Ovarian cancer (OV) involves 3% of the connected edges and Prostate cancer (PR) involves 3% of the connected edges).

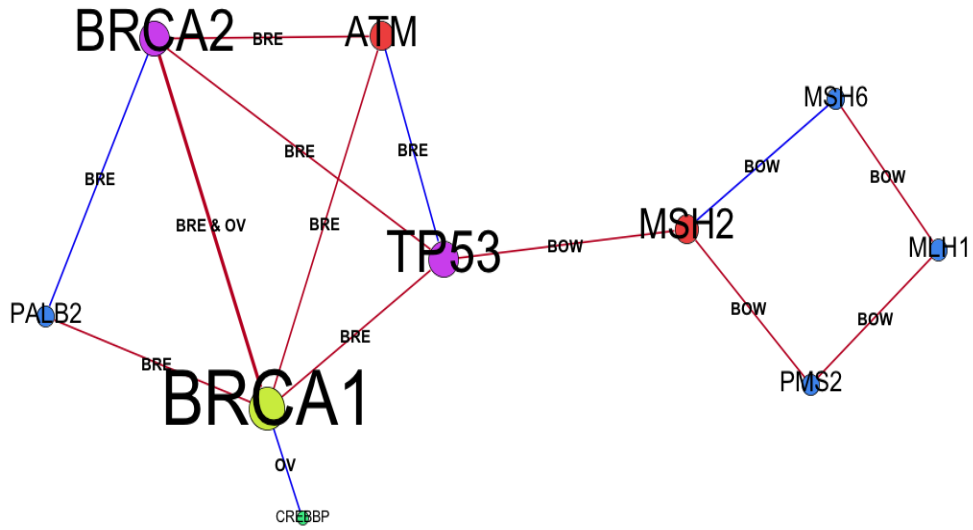


Figure 4.16: PPIM of an **Aspartate** altered amino acid and associated with the same primary cancer. 71% of the edges (red) show identical codon changes, while 29% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 50% of the connected edges, Bowel cancer (BOW) involves 36% of the connected edges and Ovarian cancer (OV) involves 14% of the connected edges).

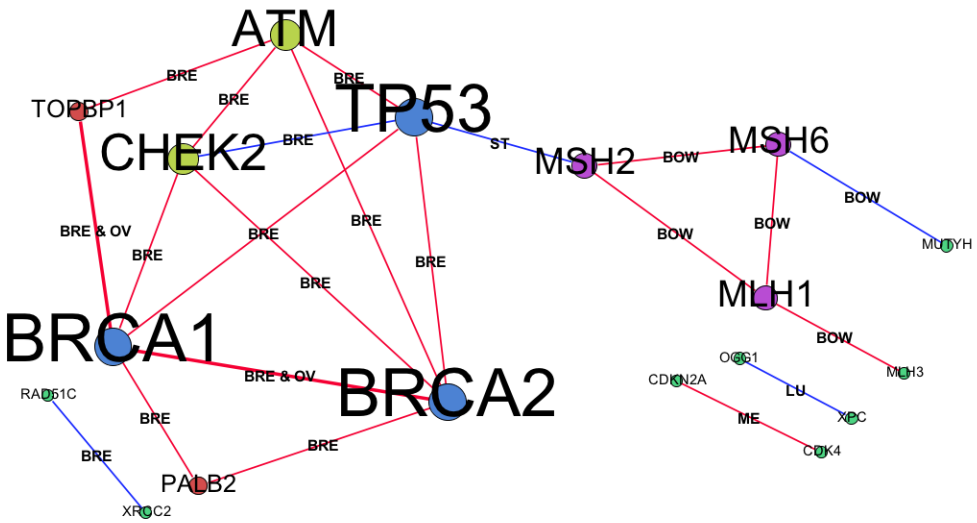


Figure 4.17: PPIM of a **Cysteine** altered amino acid and associated with the same primary cancer. 77% of the edges (red) show identical codon changes, while 23% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 53% of the connected edges, Bowel cancer (BOW) involves 23% of the connected edges, Ovarian cancer (OV) involves 9% of the connected edges, Stomach cancer (ST) involves 5% of the connected edges, Melanoma (ME) involves 5% of the connected edges and Prostate cancer (PR) involves 5% of the connected edges).

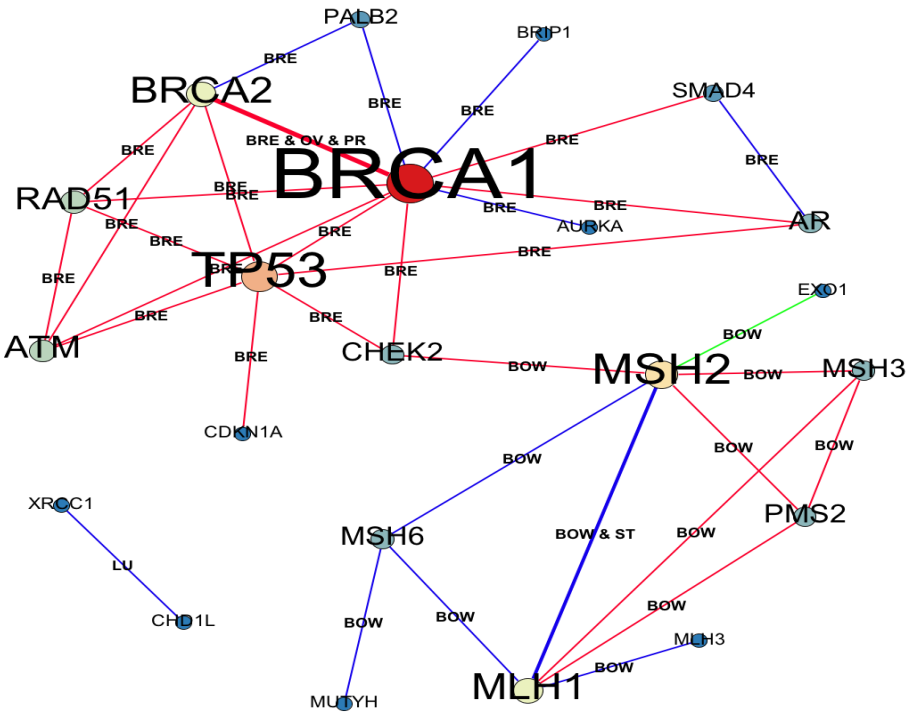


Figure 4.18: PPIM of a **Glutamine** altered amino acid and associated with the same primary cancer. 65% of the edges (red) show identical codon changes, while 32% of the edges (blue) show non-identical codon changes and the remaining 3% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 58% of the connected edges, Bowel cancer (BOW) involves 30% of the connected edges, Ovarian cancer (OV) involves 3% of the connected edges, Stomach cancer (ST) involves 3% of the connected edges and Prostate cancer (PR) involves 3% of the connected edges).

.

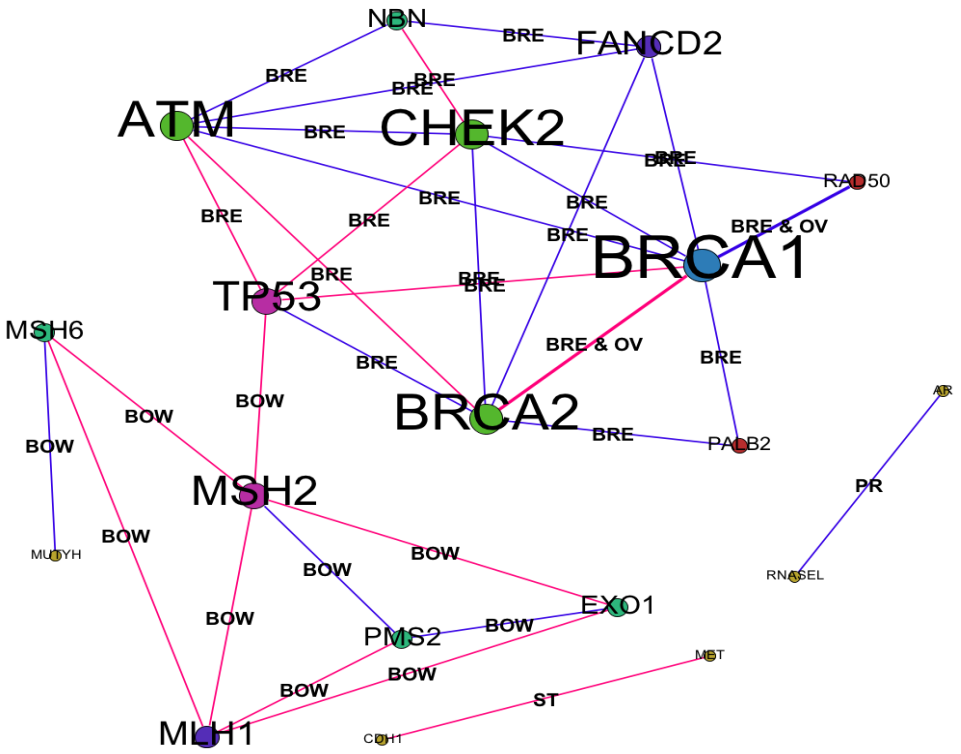


Figure 4.19: PPIM of a **Leucine** altered amino acid and associated with the same primary cancer. 44% of the edges (red) show identical codon changes, while 56% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 56% of the connected edges, Bowel cancer (BOW) involves 31% of the connected edges, Ovarian cancer (OV) involves 5% of the connected edges, Stomach cancer (ST) involves 3% of the connected edges and Prostate cancer (PR) involves 5% of the connected edges).

.

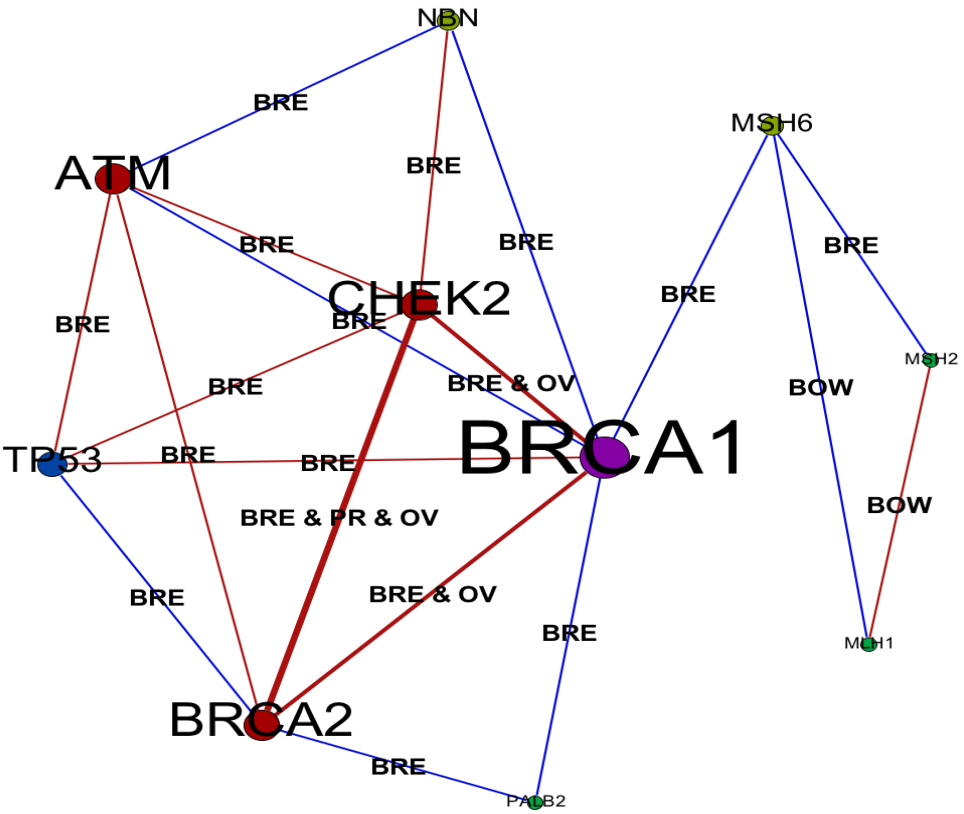


Figure 4.20: PPIM of a **Phenylalanine** altered amino acid and associated with the same primary cancer. 58% of the edges (red) show identical codon changes, while 42% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 75% of the connected edges, Bowel cancer (BOW) involves 10% of the connected edges, Ovarian cancer (OV) involves 10% of the connected edges and Prostate cancer (PR) involves 5% of the connected edges).

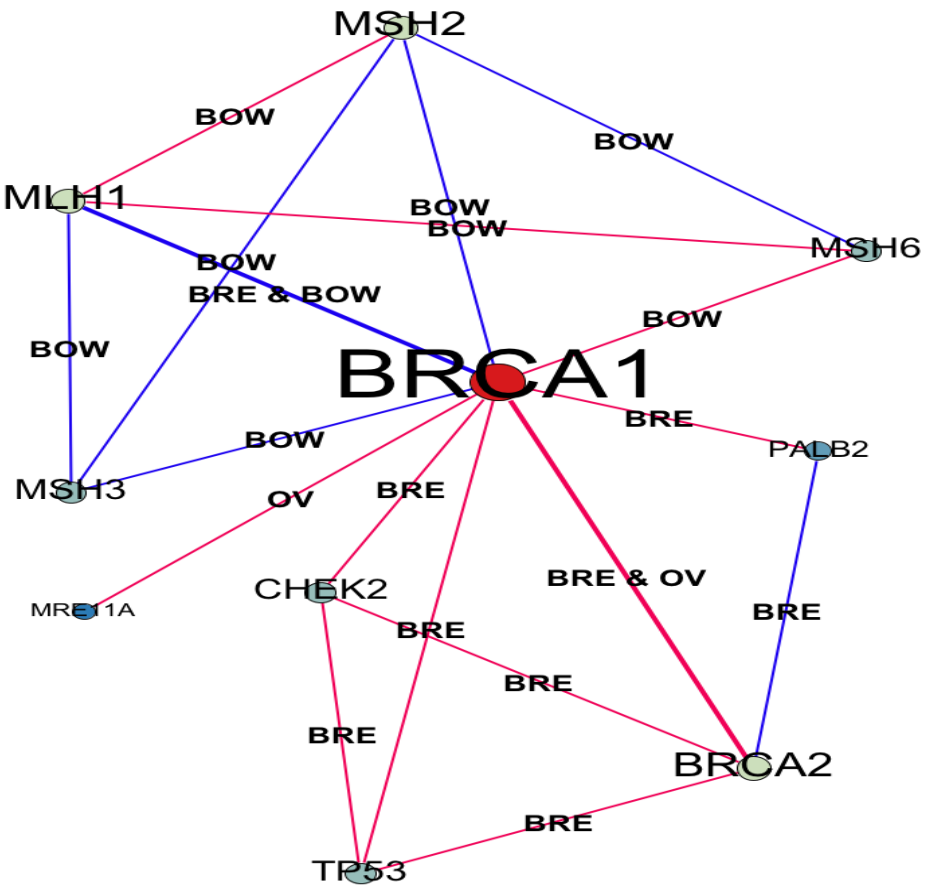


Figure 4.21: PPIM of a **Tryptophan** altered amino acid and associated with the same primary cancer. 61% of the edges (red) show identical codon changes, while 39% of the edges (blue) show non-identical codon changes. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 42% of the connected edges, Bowel cancer (BOW) involves 48% of the connected edges and Ovarian cancer (OV) involves 10% of the connected edges).

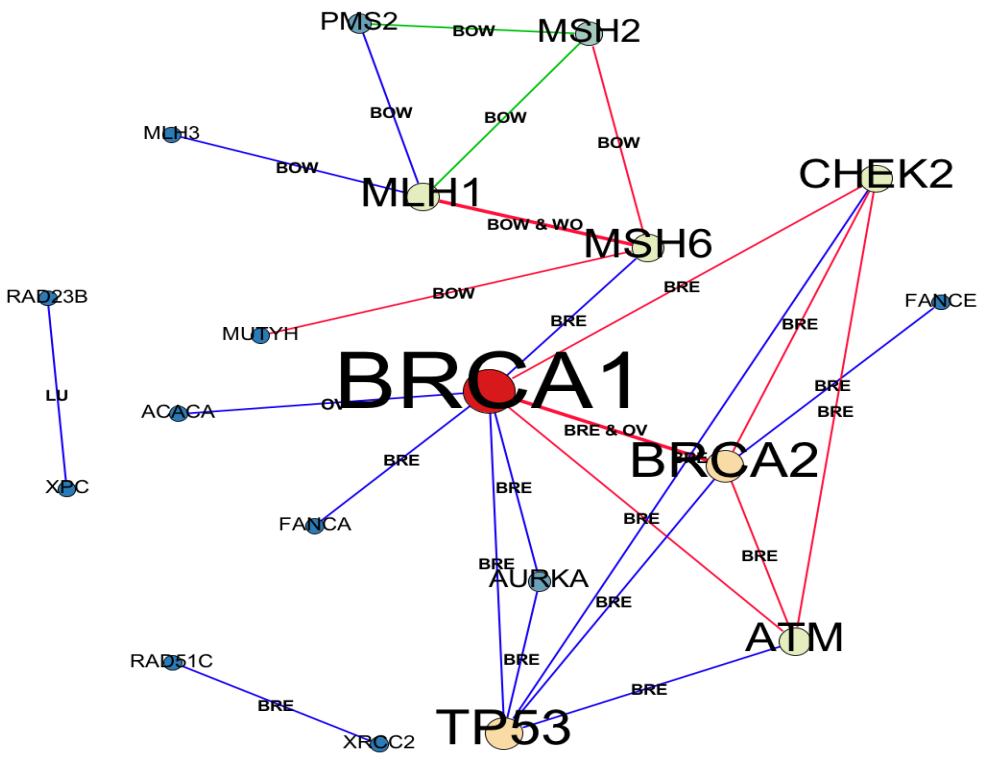


Figure 4.22: PPIM of a **Valine** altered amino acid and associated with the same primary cancer. 36% of the edges (red) show identical codon changes, while 56% of the edges (blue) show non-identical codon changes and the remaining 8% of the edges (green) show identical peptide position of the altered amino acid. The width of the edge is proportional to the number of associated cancers between the two nodes. The name on the edge is the abbreviation of the primary cancer involved (e.g, Breast cancer (BRE) involves 60% of the connected edges, Bowel cancer (BOW) involves 24% of the connected edges, Ovarian cancer (OV) involves 8% of the connected edges, Lung cancer (LU) involves 4% of the connected edges and Womb cancer (WO) involves 4% of the connected edges).

.

4.1.3. Investigating the Connectivity of the 21 Individual PPIM Altered Amino Acid Networks.

The 'altered amino acid networks' revealed several interesting points. Identical single nucleotide changes, identical altered amino acid, and identical peptide positions of the altered amino acids were found in several connected gene nodes. These data may be best presented in four parts on the basis of their sub-network connectivity. For example in the first part two fully-disconnected sub-networks occur with no bridging node, while in the second these disjointed sub-networks are bridged by *BRCA1* and *MSH6*, the third bridged by *TP53* and *MSH2*, and finally the fourth bridged by *CHEK2* and *MSH2*.

1. a) Fig. 4.2-4.14 shows two fully disconnected sub-networks, where one sub-network corresponds to connected genes associated with Bowel cancer and the other sub-network to Breast cancer. Both sub-networks also contain, a minor constituent ovarian, prostate, lung and melanoma cancers. Connected gene nodes in these networks were associated with the same amino acid alteration and/or the identical change of codon. Surprisingly some of these changes occurred at identical peptide positions in the proteins encoded by the connected genes. For example, Figure 4.7 shows that the amino acid encoded by the BRCA1 gene altered from 'Asn_{564}' to 'His_{564}', the amino acid change for the protein encoded by the *BARD1* gene was from 'Gln_{564}' to 'His_{564}'. Both amino acid alterations are associated with Breast cancer. This

finding was agreed by a previous report showing interactions of these two genes in Breast cancer (Wu et al., 1996), indicating that the *BARD1/BRCA1* interaction is interrupted by *BRCA1* Missense mutations. Others have reported that BRCA1 DNA-binding activity is stimulated by *BARD1* (Simons et al., 2006). I also detected another amino acid alteration in *BRCA1* i.e., 'Gln_{1785}' to 'His_{1785}' that was also found in the *ATM* i.e., 'Pro_{1785}' to 'His_{1785}' both associated with Breast cancer. It has been reported that DNA damage can lead to ATM-dependent phosphorylation of *BRCA1* leading to Breast cancer (Gatei et al., 2000; Cortez et al., 1999).

1.b) Fig 4.13 and 4.8 shows where *BRCA1* and *BRCA2* have an identically changed codon (i.e. 'ATG' to 'ACG') that alters the resultant amino acid from 'Met' to 'Thr' in the same peptide sequence position 1. This change is associated with Ovarian cancer. An identical codon change from 'ATG' to 'ATA' produced a 'Met' to 'Ile' alteration in same peptide also at position '1'. In this case the change was associated with Breast cancer. It has previously been reported that *BRCA1* and *BRCA2* proteins interact in pathways that lead to the activation of double-strand break repair and homologous recombination (Chen et al., 1998).

1. c) Fig 4.11 reveals an amino acid alteration 'Ala_{636}' to 'Pro_{636}' in *MSH2,* while a 'Leu_{636}' to 'Pro_{636}' amino acid alteration was apparent for *MLH1*. Both changes are associated with Bowel

cancer. Both genes also exhibit another identical amino acid alteration 'Gly_{751}' to 'Arg_{751}' in *MSH2* and 'Lys_{751}' to 'Arg_{751}' in *MLH1*, again both associated with Bowel cancer. According to previous reports these two genes are pathway genes for bowel cancer (Arends, 2013).

2. a) Fig 4.15, 4.20, 4.21 and 4.22 illustrate how two disjointed sub-networks can be connected by an identical codon or amino acid alteration. These network's 'bridges' may be a cross point for the clustering of breast and bowel cancers. For example, *BRCA1* and *MSH6* have 4 identical codon changes each resulting in an amino acid change. These include: (i) Fig 4.15 shows a 'TCT' to 'TTT' codon change resulting in an amino acid alteration from 'Ser' to 'Phe', which is associated with Breast cancer; (ii) Fig 4.20 shows a 'GCC' to 'TCC' codon change resulting in an 'Ala' to 'Ser' amino acid change also leading to Breast cancer; (iii) Fig 4.21 showing a 'CGG' to 'TGG' codon change leading to a 'Arg' to 'Trp' amino acid change that in this case is associated Bowel cancer. Fig 4.22 shows a change in amino acid from 'Ile' to 'Val' for *BRCA1* and *MSH6* that is associated with Breast cancer.

BRCA1 is connected with two other genes (MSH2 and MLH1) as they are associated with Bowel cancer. The literature reports that *MSH6, MSH2* and *MLH1* genes are associated with *BRCA1* to form a large protein complex for the recognition of abnormal/damaged DNA structures (Wang et al., 2000).

3) Fig 4.16, 4.17 and 4.19 reveal networks with two disjoined sub-networks that are bridged with a *TP53* and *MSH2* connection. Both genes share an identical codon change that produces different amino acids. In Fig 4.16 this is 'GGC' to 'GAC' resulting in 'Gly' to 'Asp' switch, while in Fig 4.19 the change of codon is from 'CCC' to 'CTC', with resultant 'Pro' to 'Leu' switch. Both changes are associated with Bowel cancer. By contrast, Fig 4.17 shows another sub-network linked this time by *TP53* and *MSH2.* The encoded protein TP53 is altered via the amino acid change 'Arg' to 'Cys', while for MSH2 'Tyr' is altered to 'Cys'; both changes are associated with Stomach cancer. The literature reports that these two genes bind together during the DNA synthesis phase of the cell cycle (Zink et al., 2002).

4) Fig 4.18 shows an identical codon change in *CHEK2* and *MSH2* that connect the two disjoined sub-networks. Here the codon has changed from 'CGG' to 'CAG' in both genes, resulting in a 'Arg' to 'Gln' switch and is associated with Bowel cancer. The literature reports that *MSH2* binds *CHEK2* (Brown et al., 2002).

4.2. Discussion/Conclusions

The main objectives of this chapter were as follows: (i) associating genes based on the physical interactions of their encoded proteins and thereby construct a protein-protein interaction map (PPIM), (ii) extracting Missense/Nonsense mutations in codons and determine their associated amino acid changes, and relating these to associated cancers within the PPIM, (iii) merging the extracted Missense/Nonsense mutation records with the PPIM data to construct “21 amino acid alteration networks”, and lastly; (iv) investigating the influence of a single nucleotide change, amino acid change and the peptide position of the changed amino acid in the altered amino acid networks.

The amino acid alteration networks revealed:

1. Mutation-derived alterations in Alanine, Arginine, Asparagine, Glutamate, Histidine, Isoleucine, Lysine, Methionine, Proline, Termination, Threonine and Tyrosine were associated with two fully-disjoint sub-networks; the first sub-network associated with genes linked to Bowel cancer and the second sub-network associated with genes leading to Breast cancer.
2. Mutation-derived alterations in Serine, Aspartate, Cysteine, Glutamine, Leucine, Phenylalanine, Tryptophan and Valine connected the two disjoint sub-networks and linked the clusters of Bowel, Breast or Stomach cancers.

3. *BRCA1* and *MSH6* appear to play a central role in connecting the disjoint sub-networks of mutation-derived alterations in Serine, Phenylalanine, Tryptophan and Valine, where both genes are associated with Breast and Bowel cancers.
4. *TP53* and *MSH2* connect disjointed sub-networks by the encoded amino acid alterations in Leucine and Aspartate, resulting in Bowel cancer, while the encoded amino acid alteration leading to Cysteine was associated with Stomach cancer.
5. *CHEK2* and *MSH2* genes connect genes in the disjointed sub-networks, but the encoded amino acid change to Glutamine, resulting in Bowel cancer.
6. Identical encoded amino acid change at identical peptide sequence positions for two associated genes leading to same cancer.

It may be useful to further investigate the groups of associated genes in these altered amino acid networks, and especially those gene nodes that connect disjointed gene node clusters in the network. Areas of potential relevant research could be around their enzyme binding, gene ontology, damaged DNA binding and DNA repair etc.

Chapter Five

5. Overall Discussion and Future Work

5.1. General Conclusions

The aim of this project was to assess the involvement of germline mutation classes in developing human cancers. This aim was progressed via several specific objectives/questions as follows: (i) Are individual groups of primary cancers associated with specific gene subsets groups (within the 424 cancer genes)? (ii) Are there any specific groups of primary cancers associated with particular mutation classes? (iii) If both questions prove true, are groups of cancers associated with particular mutation classes of target gene groups? This project also explored whether a corresponding Protein-Protein Interaction Map, derived from the Missense/Non-sense Mutation portion of the HCM gene set, would provide further information on gene associations between primary cancers in terms of the consequent identical amino acid changes involved.

Results showed that: (1) Closely-connected human cancers in the HCM exhibited a strong association with particular mutation class. (2) Missense /Nonsense and Regulatory mutations played a central role in

connecting cancers (i.e. via primary cancer nodes) and so significantly influenced the construction of the HCM. (3) Group of genes with Missense/Nonsense and Regulatory mutations tended to be cancer-associated pathway genes. (4) *BRCA1, BRCA2, PALB2, MSH2, MSH6, MLH1, CDKN2A,* and *TP53* showed the highest agreement for 5 of 10 mutation classes. (5) From the PPIM, it was evident that *BRCA1, MSH6, BARD1, TP53, MSH2* and *CHEK2* proteins best connected Breast, Ovarian, Prostate and Bowel primary cancers and so could represent 'driver proteins' for these cancers.

In summary, this project has approached the analysis of gene involvement in human primary cancers from the starting position of the mutation class that harbours the specific gene mutation. Together with their downstream resultant alterations in the associated proteins, this analysis provides great insights into the relatedness of primary human cancers and their associated potential gene hierarchies. The data produced in this study may therefore help us to understand further the aetiology, diagnosis and potentially personalized treatments for cancer. For example, the amino acids connecting the disjoined networks via different genes and cancers at the missense/nonsense mutation might be a point where one gene is affected leading to an effect on all linked genes in the networks, which could then lead to the development of other cancers.

5.2. General Discussion

The Human Cancer Map (HCM) revealed a significant variation in the dominance of different mutation class involved in cancer. Missense/Nonsense mutation (44%) represented almost half of the interconnections in the HCM, followed by small deletion mutations at 14%. The remaining mutation classes ranged from 10% to 1%. This means that cancer associations vary from one mutation class to another. For example, a total of 19 primary cancers with Missense/Nonsense mutation were found to be associated to each other, whereas only 12 primary cancers were represented by splicing mutations.

The agreements between the connected nodes (cancers) in the HCM and sub-HCMs (i.e. sub-MAPs for each mutation class) were examined and quantified using Choen's kappa test (Carletta, 1996; Cohen, 1960), and revealed some interesting results. Cancer node connectivity levels involving Missense/Nonsense and the Regulatory mutation classes ranged from low for 80% of the nodes ($k<0.25$) to moderate for 20% ($k=0.25$-0.5). This suggests that although this mutation class dominates the HCM, it may be less important mutations in terms of connectivity between cancer nodes in the constructed HCM (Chapter 3, Figure 3.13). However, many more of these genes are involved in pathways (KEGG) for cancers and other disorders than for other mutation classes, even if the later show higher levels of kappa agreement (Chapter 3, Figure 3.26 and Table 3.1). By contrast, mutations involving Repeat variations, Complex rearrangement and Gross insertions showed low

agreement values for all the connected nodes, which make them less important mutations in terms of connectivity between cancer nodes in the constructed HCM. (Chapter 3, Figure 3.13 and Appendix 1 Tables A.1.8, A.1.9 and A.1.10). Furthermore, these gene nodes were not represented in the KEGG pathways for cancer. Surprisingly, while the remaining five mutation classes showed some high kappa agreement values between interconnected gene nodes, the mutation class was represented in only a small proportion of all nodes (Fig. 3.13). For example, while Small Indels mutations represented only a mere 2% of the total HCM, their corresponding inter-node connectivity revealed 100% high or very high kappa values (agreement) (Chapter 3, Figure 3.13 and Appendix 1 Tables A.1.2, A.1.3, A.1.4, A.1.6 and A.1.7).

From a biological perspective, what does the above finding mean? I would like to suggest two potential hypotheses. The first indicates that a Missense/Nonsense and Regulatory mutation has a large impact on how cancer nodes associate with each other in a HCM. Without the involvement of these two mutation classes the HCM would fragment, and almost 50% of cancer nodes would be disconnected. This then may suggest that mutations of these 2 classes may contain driver mutations connecting cancer gene nodes. Alternatively, mutations of the following types (i.e. Small Deletions, Splicing, Small Insertions, Gross Deletions, and Small Indels), with their high to very high agreements (kappa values) between some of their connected cancer nodes, may make them more deleterious as they would have greater impact on cancer gene

connectivity in the HCM. For example, Small deletion mutations tend to be a target mutation connecting Womb cancer to Lymphoma and Brain cancer nodes via small deletion mutations in the *BRCA2, BRCA1* and *PALB2* genes, while at the same time connecting Lymphoma to Brain cancer nodes via small deletion mutations in *MSH2, MSH6, BRCA1* and *MLH1* (figure 3.3 and Appendix 1 Table A1.2). Splicing mutations tend to be target mutations for connecting Melanoma to Brain and Pancreatic cancer nodes via splicing mutations in the *CDKN2A* gene, and for connecting Breast to Ovarian cancer nodes via splicing mutations in *BRCA1, BRCA2, RAD51C, RAD51D* and *TP53* genes. They also connect Kidney to Bladder cancer nodes via splicing mutations in *MSH2* (Figure 3.4 and Appendix 1 Table A1.3).

Small Insertion mutations tend to be more engaged in the HCM by targeting more high to very high agreements (kappa values) between their inter-connected cancer nodes. For example, Stomach cancer connected to Bowel and Kidney cancers via small Insertion mutations in *CDH1 and MSH2*, while Breast cancer connected to Ovarian cancer via small Insertion mutations *BRCA1, BRIP1* and *RAD51C*. Melanoma was connected to Lung, Prostate and Pancreatic cancers via small Insertion mutations of *BRCA2* and *CDKN2A*, while Pancreatic cancer was connect to Prostate and Lung by small Insertion mutations in *BRCA2* (Figure 3.5 and Appendix 1 Table A1.4).

Finally, Small Indels mutation tend to have a high to very high agreements (100%) between all its connected cancer nodes, such as

Bowel connected to the Brain and Womb cancer nodes via *MLH1 and MSH6*, Ovarian connected to Breast and Pancreatic cancer node via *BRCA1 and BRCA2*, Melanoma connected to the Head and Neck cancer node via *CDKN2A* (Figure 3.8 and Appendix 1 Table A1.7).

The above five mutation classes may share some similarities in terms of their genetics implications via associated mutations in *BRCA1, BRCA1, PALB2, MSH6, MSH2, MLH1, CDKN2A, RAD51C, RAD51D, TP53, CDH1 and BRIP1*. These 12 genes were also connected in the Genome-wide Distribution Maps and also at the protein interaction as shown in the Protein-Protein Interactions Map (PPI). DAVID analysis revealed that 2 of these twelve genes (i.e., *MSH6* and *MLH1*) tended to have similar biological functions. All these revealed information suggesting these groups of genes could be the main driver genes, which are responsible for many types of cancers as shown in this study. These relations can be seen in Chapter 3, figures 3.16 to 3.24 and Chapter 4 figure 4.1.

The results of this study suggested that Missense/Nonsense mutations have a large influence in interconnecting cancer nodes, and interrogation of these Missense/Nonsense mutation records provided new insights. One of these is that cancers caused by different amino acid alterations involving all 21 amino acids of the human body. The interrogation results targeted two groups of interconnected genes; each one formed a network associated with a specific cancer. The first group of connected genes (*MSH2, PMS2, MLH1, MSH6* and *MUTYH*) associated

with Bowel cancer, while the second group of genes (*PALB2, BRCA1, BRCA2, ATM, TP53* and *CHEK2*) associated with Breast cancer. These PPIM networks occur when their nucleotide changes encode alterations in one of the following 13 amino acids (i.e., Alanine, Arginine, Asparagine, Glutamate, Glycine, Histidine, Isoleucine, Lysine, Methionine, Proline, Termination, Threonine and Tyrosine) (Chapter 4 figures 4.2 to 4.14). But when these two groups of genes undergo nucleotide change to encode one of the following 8 amino acids (i.e., Serine, Aspartate, Cysteine, Glutamine, Leucine, Phenylalanine, Tryptophan and Valine) we see merging of the two disjointed node clusters in the network. The bridging is either by Breast, Bowel or Ovarian cancer (figures 4.15 to 4.22), and could be a interesting subject for further research as these 12 genes appeared to be significant throughout this study, ie are over represented in cancers.

5.3. Future Work

It would be potentially very useful to extending the current study presented in this thesis by considering the following important points:

1. Extending current work on the HCM, which currently includes 20 primary human cancers and their associated 424 genes (of a total of 29 primary human cancers), by including updated information of genes, cancers and mutation records. This will allow us to produce a global HCM.
2. Extending the current interrogation protocol used for Missense/Nonsense mutation class to the remaining nine mutation classes i.e. Splicing, Small deletions, Small Insertion, Small Indels, Complex Rearrangement, Gross deletion, Gross Insertion, Regulatory and Repeat variation.
3. Investigating the environmental factors that have impact on gene behaviour, which will address unanswered questions in relation to genetic mutations. This will allow us to produce a modified Global Human Cancer Map.
4. Integrating somatic cell mutations and germline mutations.
5. Investigating further the following highly represented cancer-related genes i.e. *MSH2, PMS2, MLH1, MSH6, MUTYH, PALB2, BRCA1, BRCA2, ATM, TP53 and CHEK2*) in terms of their associated biological activities e.g., enzyme binding, gene ontology, Damage DNA binding and DNA repair etc.

6. Expands the current work presented in chapter three (figure 3.15), to include a patient based clinical data and increase to the number of gene, mutation and cancer records.

References

Ahvenainen T, Lehtonen HJ, Lehtonen R, et al. (2008) Mutation screening of fumarate hydratase by multiplex ligation-dependent probe amplification: detection of exonic deletion in a patient with leiomyomatosis and renal cell cancer. Cancer Genetics and Cytogenetics 183: 83-88.

Bardella C, El-Bahrawy M, Frizzell N, et al. (2011) Aberrant succination of proteins in fumarate hydratase-deficient mice and HLRCC patients is a robust biomarker of mutation status. The Journal of Pathology 225: 4-11.

Bonadona V, Bonaïti B, Olschwang S, et al. (2011) Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA: The Journal of the American Medical Association 305: 2304.

Brown TA. (1992) Genetics: a molecular approach: Chapman & Hall.

Bruder CEG, Piotrowski A, Gijsbers AACJ, et al. (2008) Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. The American Journal of Human Genetics 82: 763-771.

Brunet J, Gutiérrez-Enríquez S, Torres A, et al. (2008) ATM germline mutations in Spanish early-onset breast cancer patients negative for BRCA1/BRCA2 mutations. Clinical Genetics 73: 465-473.

Cooper DN, Chen JM, Ball EV, et al. (2010) Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. Human Mutation 31: 631-655.

Del-Favero J, Goossens D, Van den Bossche D, et al. (1999) YAC fragmentation with repetitive and single-copy sequences: detailed physical mapping of the presenilin 1 gene on chromosome 14. Gene 229: 193-201.

Engert S, Wappenschmidt B, Betz B, et al. (2008) MLPA screening in the BRCA1 gene from 1,506 German hereditary breast cancer cases: novel deletions, frequent involvement of exon 17, and occurrence in single early-onset cases. Human Mutation 29: 948-958.

Epstein DJ. (2009) Cis-regulatory mutations in human disease. Briefings in Functional Genomics & Proteomics 8: 310.

Flotte TR and Laube BL. (2001) Gene therapy in cystic fibrosis. CHEST Journal 120: 124S-131S.

Ford D, Easton D, Stratton M, et al. (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. American Journal of Human Genetics 62: 676.

Griffiths AJ. (2005) An introduction to genetic analysis: New York: WH Freeman.

Hames BD and Hooper NM. (2000) Instant notes in biochemistry: Bios Scientific Pub Ltd.

Hill KA, Wang J, Farwell KD, et al. (2003) Spontaneous tandem-base mutations (TBM) show dramatic tissue, age, pattern and spectrum specificity. Mutation Research/Genetic Toxicology and Environmental Mutagenesis 534: 173-186.

Jimenez-Sanchez G, Childs B and Valle D. (2001) Human disease genes. Nature 409: 853-855.

Kim YM, Cho SK and Chang WY. (2010) Three novel germline mutations in MLH1 and MSH2 in families with Lynch syndrome living on Jeju island, Korea. Biochemistry and Molecurar Biology Reports 43: 693-697.

Kimball JW. (2003) Kimball's biology pages: John W. Kimball.

Kondrashov AS and Rogozin IB. (2004) Context of deletions and insertions in human coding sequences. Human Mutation 23: 177-185.

Kwon JA, Lee SY, Ahn EK, et al. (2010) Short rare MUC6 minisatellites-5 alleles influence susceptibility to gastric carcinoma by regulating gene. Human Mutation 31: 942-949.

La Paglia L, Laugé A, Weber J, et al. (2010) ATM germline mutations in women with familial breast cancer and a relative with haematological malignancy. Breast Cancer Research and Treatment 119: 443-452.

Lenaers G, Hamel C, Delettre C, et al. (2012) Dominant optic atrophy. Orphanet Journal of Rare Diseases 7: 1-12.

Lodish H, Berk A, Zipursky SL, et al. (2000) Molecular cell biology. New York.

Meuwissen ME and Mancini G. (2012) Neurological findings in incontinentia pigmenti; a review. European Journal of Medical Genetics.

MLS DHC and MLIS AMH. (2005) Genetics Home Reference. Journal of Consumer Health on the Internet 9: 61-68.

Moat SJ, Bradley DM, Salmon R, et al. (2013) Newborn bloodspot screening for Duchenne muscular dystrophy: 21 years experience in Wales (UK). European Journal of Human Genetics.

Nilbert M, Wikman FP, Hansen TV, et al. (2009) Major contribution from recurrent alterations and MSH6 mutations in the Danish Lynch syndrome population. Familial Cancer 8: 75-83.

Norvell JC and Machalek AZ. (2000) Structural genomics programs at the US national institute of general medical sciences. Nature Structural & Molecular Biology 7: 931-931.

Ogurtsov AY, Sunyaev S and Kondrashov AS. (2004) Indel-based evolutionary distance and mouse–human divergence. Genome Research 14: 1610.

Oti M, Huynen MA and Brunner HG. (2008) Phenome connections. Trends in Genetics 24: 103-106.

Parry EM. (2001) Human Chromosomes (4th edn). Heredity 87: 508-509.

Pasternak JJ. (2005) An introduction to human molecular genetics: mechanisms of inherited diseases: Wiley-Blackwell.

Qiu C, Zhou W and Song C. (1996) [Nucleotide sequence of HLA-DQA1 promoter region (QAP) in a lung cancer patient]. Zhongguo yi xue ke xue yuan xue bao. Acta Academiae Medicinae Sinicae 18: 172.

Ramus SJ, Harrington PA, Pye C, et al. (2007) Contribution of BRCA1 and BRCA2 mutations to inherited ovarian cancer. Human Mutation 28: 1207-1215.

Roberts JAF and Pembrey ME. (1970) An introduction to medical genetics: Oxford University Press Oxford.

Sibley K, Rollinson S, Allan JM, et al. (2003) Functional FAS promoter polymorphisms are associated with increased risk of acute myeloid leukemia. Cancer Research 63: 4327-4330.

Sluiter MD and van Rensburg EJ. (2011) Large genomic rearrangements of the BRCA1 and BRCA2 genes: review of the literature and report of a novel BRCA1 mutation. Breast Cancer Research and Treatment 125: 325-349.

Sweet K, Senter L, Pilarski R, et al. (2010) Characterization of BRCA1 ring finger variants of uncertain significance. Breast Cancer Research and Treatment 119: 737-743.

Vaughn CP, Robles J, Swensen JJ, et al. (2010) Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Human Mutation 31: 588-593.

Walsh T, Lee MK, Casadei S, et al. (2010) Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proceedings of the National Academy of Sciences 107: 12629-12633.

Worsley-Hunt R, Bernard V and Wasserman WW. (2011) Identification of cis-regulatory sequence variations in individual genome sequences. Genome Medicine 3: 1-14.

Yamakawa-Kobayashi K, Kobayashi T, Yanagi H, et al. (1994) A novel complex mutation in the LDL receptor gene probably caused by the

simultaneous occurrence of deletion and insertion in the same region. Human Genetics 93: 625-628.

Yang M, Guo H, Wu C, et al. (2009) Functional FEN1 polymorphisms are associated with DNA damage levels and lung cancer risk. Human Mutation 30: 1320-1328.

Yang Y and Zhang S. (2010) Y chromosome variations and male infertility. Chinese Journal of Medical Genetics 27: 276.

Ahmadian A, Gharizadeh B, Gustafsson AC, et al. (2000) Single-nucleotide polymorphism analysis by pyrosequencing. Analytical biochemistry 280: 103-110.

Amberger J, Bocchini CA, Scott AF, et al. (2009) McKusick's Online Mendelian Inheritance in Man (OMIM®). Nucleic acids research 37: D793.

Arends MJ. (2013) Pathways of Colorectal Carcinogenesis. Applied Immunohistochemistry & Molecular Morphology 21: 97-102.

Barabási AL, Gulbahce N and Loscalzo J. (2011) Network medicine: a network-based approach to human disease. Nature Reviews Genetics 12: 56-68.

Bastian M, Heymann S and Jacomy M. (2009) Gephi: An open source software for exploring and manipulating networks. International AAAI conference on weblogs and social media. AAAI Press Menlo Park, CA.

Becker KG, Barnes KC, Bright TJ, et al. (2004) The genetic association database. Nature genetics 36: 431-432.

Bell SP, Gann A, Levine M, et al. (2003) Molecular Biology of the Gene. Benjamin-Cummings. Nueva Cork.

Boutet E, Lieberherr D, Tognolli M, et al. (2007) Uniprotkb/swiss-prot. Plant Bioinformatics. Springer, 89-112.

Brown KD, Rathi A, Kamath R, et al. (2002) The mismatch repair system is required for S-phase checkpoint activation. Nature genetics 33: 80-84.

Cabanillas R, Astudillo A, Valle M, et al. (2011) Novel germline CDKN2A mutation associated with head and neck squamous cell carcinomas and melanomas. Head & Neck.

Cardon LR, Craddock N, Deloukas P, et al. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661-678.

Carletta J. (1996) Assessing agreement on classification tasks: the kappa statistic. Computational linguistics 22: 249-254.

Chaudhuri S and Dayal U. (1997) An overview of data warehousing and OLAP technology. ACM Sigmod record 26: 65-74.

Chen J, Silver DP, Walpita D, et al. (1998) Stable interaction between the products of the BRCA1 and BRCA2 tumor suppressor genes in mitotic and meiotic cells. Molecular cell 2: 317.

Chu S, DeRisi J, Eisen M, et al. (1998) The transcriptional program of sporulation in budding yeast. Science 282: 699-705.

Cohen J. (1960) A coefficient of agreement for nominal scales. Educational and psychological measurement 20: 37-46.

Cohen J. (1968) Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. Psychological bulletin 70: 213.

Conrad DF, Keebler JE, DePristo MA, et al. (2011) Variation in genome-wide mutation rates within and between human families. Nature 201: 1.

Cortez D, Wang Y, Qin J, et al. (1999) Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. Science 286: 1162-1166.

Da Wei Huang BTS and Lempicki RA. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols 4: 44-57.

Dagley S and Nicholson DE. (1970) An introduction to metabolic pathways. An introduction to metabolic pathways.

Davies H, Bignell GR, Cox C, et al. (2002) Mutations of the BRAF gene in human cancer. Nature 417: 949-954.

de Snoo FA, Bishop DT, Bergman W, et al. (2008) Increased risk of cancer other than melanoma in CDKN2A founder mutation (p16-Leiden)-positive melanoma families. Clinical Cancer Research 14: 7151-7157.

Dreze M, Monachello D, Lurin C, et al. (2010) High-quality binary interactome mapping. Methods in enzymology 470: 281-315.

Futreal PA, Coin L, Marshall M, et al. (2004) A census of human cancer genes. Nature Reviews Cancer 4: 177-183.

Gatei M, Scott SP, Filippovitch I, et al. (2000) Role for ATM in DNA damage-induced phosphorylation of BRCA1. Cancer research 60: 3299-3304.

Goehler H, Lalowski M, Stelzl U, et al. (2004) A protein interaction network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. Molecular cell 15: 853-865.

Goh K-I, Cusick ME, Valle D, et al. (2007) The human disease network. Proceedings of the National Academy of Sciences 104: 8685-8690.

Goldstein AM. (2004) Familial melanoma, pancreatic cancer and germline CDKN2A mutations. Human mutation 23: 630-630.

Grzybowska E, Siemiñska M, Zientek H, et al. (2002) Germline mutations in the BRCA1 gene predisposing to breast and ovarian cancers in Upper Silesia population. ACTA BIOCHIMICA POLONICA-ENGLISH EDITION- 49: 351-356.

Guelzim N, Bottani S, Bourgine P, et al. (2002) Topological and causal structure of the yeast transcriptional regulatory network. Nature genetics 31: 60-63.

Halevy AY. (2001) Answering queries using views: A survey. The VLDB Journal—The International Journal on Very Large Data Bases 10: 270-294.

Jensen UB, Sunde L, Timshel S, et al. (2010) Mismatch repair defective breast cancer in the hereditary nonpolyposis colorectal cancer syndrome. Breast cancer research and treatment 120: 777-782.

Johnson A and O'Donnell C. (2009) An open access database of genome-wide association results. BMC medical genetics 10: 6.

Joshi A. (1997) The Language of the Genes. RESONANCE: 90.

Kanehisa M, Goto S, Hattori M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic acids research 34: D354-D357.

Kinzler KW and Vogelstein B. (1998) The genetic basis of human cancer: McGraw-Hill Toronto:.

Koch GG. (1983) Intraclass correlation coefficient. Encyclopedia of statistical sciences.

Kozak M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell 44: 283-292.

Krzywinski M, Schein J, Birol İ, et al. (2009) Circos: an information aesthetic for comparative genomics. Genome Research 19: 1639-1645.

Leja N. (2010) Chromosome. National Human Genome Research Institute: genome.gov.

Mardis ER, Ding L, Dooling DJ, et al. (2009) Recurring mutations found by sequencing an acute myeloid leukemia genome. New England Journal of Medicine 361: 1058-1066.

Medicine USNLo. (2013a) Deletion Mutation. Genetic Home Reference.

Medicine USNLo. (2013b) Insertion Mutation. Genetic Home Reference.

Medicine USNLo. (2013c) Missense. Genetic Home Reference.

Mendel G. (2008) Experiments in plant hybridisation: Cosimo, Inc.

Montojo J, Zuberi K, Rodriguez H, et al. (2010) GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics 26: 2927-2928.

Nanjundan M, Nakayama Y, Cheng KW, et al. (2007) Amplification of MDS1/EVI1 and EVI1, located in the 3q26. 2 amplicon, is associated with favorable patient prognosis in ovarian cancer. Cancer research 67: 3074-3084.

Newman ME. (2003) The structure and function of complex networks. SIAM Review 45: 167-256.

Novak K. (2004) Epigenetics Changes in Cancer Cells: Highlights of the American Association for Cancer Research Special Conference on Chromatin, Chromosomes, and Cancer Epigenetics; November 10-14, 2004; Waikoloa, Hawaii. Medscape General Medicine 6.

Proulx SR, Promislow DE and Phillips PC. (2005) Network thinking in ecology and evolution. Trends in Ecology & Evolution 20: 345-353.

Quackenbush J. (2011) The human genome: Charlesbridge Publishing.

Rebhan M, Chalifa-Caspi V, Prilusky J, et al. (1997) GeneCards: integrating information about genes, proteins and diseases. Trends in genetics: TIG 13: 163.

Rosenfeld P, McKusick V, Amberger J, et al. (1994) Recent advances in the gene map of inherited eye disorders: primary hereditary diseases of the retina, choroid, and vitreous. Journal of medical genetics 31: 903.

S Dole DC, T. Wright & G. Hilton. (2009) Exploring mass, volume & density. The University of Queensland, Proportional Reasoning 1.

Safran M, Solomon I, Shmueli O, et al. (2002a) GeneCards™ 2002: towards a complete, object-oriented, human gene compendium. Bioinformatics 18: 1542.

Safran M, Solomon I, Shmueli O, et al. (2002b) GeneCards™ 2002: towards a complete, object-oriented, human gene compendium. Bioinformatics 18: 1542-1543.

Sherman BT, Tan Q, Kir J, et al. (2007) DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic acids research 35: W169-W175.

Simons AM, Horwitz AA, Starita LM, et al. (2006) BRCA1 DNA-binding activity is stimulated by BARD1. Cancer research 66: 2012-2018.

Stark C, Breitkreutz B-J, Chatr-Aryamontri A, et al. (2011) The BioGRID interaction database: 2011 update. Nucleic acids research 39: D698-D704.

Stenson PD, Mort M, Ball EV, et al. (2009) The human gene mutation database: 2008 update. Genome Med 1: 13.

Stratton MR, Campbell PJ and Futreal PA. (2009) The cancer genome. Nature 458: 719-724.

Wagner A. (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. Molecular Biology and Evolution 18: 1283-1292.

Wagner A and Fell DA. (2001) The small world inside large metabolic networks. Proceedings of the Royal Society of London. Series B: Biological Sciences 268: 1803-1810.

Wang Y, Cortez D, Yazdi P, et al. (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes & development 14: 927-939.

Welch EM, Barton ER, Zhuo J, et al. (2007) PTC124 targets genetic disorders caused by nonsense mutations. Nature 447: 87-91.

Welcsh PL and King M-C. (2001) BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. Human Molecular Genetics 10: 705-713.

Whelan AJ, Bartsch D and Goodfellow PJ. (1995) A familial syndrome of pancreatic cancer and melanoma with a mutation in the CDKN2 tumor-suppressor gene. New England Journal of Medicine 333: 975-977.

Wray GA. (2007) The evolutionary significance of cis-regulatory mutations. Nature Reviews Genetics 8: 206-216.

Wu LC, Wang ZW, Tsan JT, et al. (1996) Identification of a RING protein that can interact in vivo with the BRCA1 gene product. Nature genetics 14: 430-440.

Zink D, Mayr C, Janz C, et al. (2002) Association of p53 and MSH2 with recombinative repair complexes during S phase. Oncogene 21: 4788-4800.

Appendix

Appendix 1

Table A.1.1: Missense/Nonsense SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 56 genes and 19 primary cancer nodes.

Cancer Node1	Cancer Node 2	Shared Genes	Kappa
Lung Cancer	Prostate Cancer	CYP1B1	-0.20
Bowel Cancer	Breast Cancer	NBN,TP53,MSH6,PML,BRCA 1,BRCA2,CHEK2,PARP1	-0.17
Bowel Cancer	Ovarian Cancer	TP53,BRCA1,BRCA2,CHEK2	-0.14
Lung Cancer	Bowel Cancer	ABCB1,AXIN2,CYP1A1,NBN, OGG1,TP53,XPC	-0.14
Lung Cancer	Ovarian Cancer	BARD1,TNFRSF10A,TP53	-0.14
Lung Cancer	Melanoma	NBN	-0.14
Bowel Cancer	Pancreatic Cancer	BRCA2	-0.13
Lung Cancer	Breast Cancer	ATM,BARD1,EGFR,ERBB2,N BN,TP53,XRCC1	-0.09
Breast Cancer	Melanoma	NBN,BRCA2	-0.07
Lung Cancer	Stomach Cancer	TP53	-0.07
Melanoma	Ovarian Cancer	BRCA2	-0.07
Bowel Cancer	Skin Cancer (Non Melanoma)	TP53	-0.06
Bowel Cancer	Kidney Cancer	FLCN,MSH2	-0.05
Kidney Cancer	Ovarian Cancer	GEMIN4	-0.05
Skin Cancer(Non Melanoma)	Breast Cancer	TP53	-0.05
Breast Cancer	Pancreatic Cancer	BRCA2,FANCA	-0.04
Bowel Cancer	Brain tumours	TP53,MSH2	-0.02
Breast Cancer	Head and neck cancer	EGFR	-0.02
Breast Cancer	Oesophageal Cancer	AURKA	-0.02
Melanoma	Prostate Cancer	BRCA2	-0.02
Ovarian Cancer	Brain tumours	TP53	-0.02
Womb Cancer	Ovarian Cancer	BRCA1	-0.02
Bowel Cancer	Melanoma	NBN,BMP4,BRCA2	-0.01
Breast Cancer	Stomach Cancer	TP53,CDH1	0.00
Breast Cancer	Brain tumours	TP53,PTPN11	0.00
Breast Cancer	Womb Cancer	MSH6,BRCA1	0.00

Prostate Cancer	Pancreatic Cancer	BRCA2	0.01
Skin Cancer(Non Melanoma)	Ovarian Cancer	TP53	0.01
Lung Cancer	Brain tumours	TP53,BAP1	0.02
Stomach Cancer	Prostate Cancer	CDH1	0.03
Womb Cancer	Prostate Cancer	BRCA1	0.03
Bowel Cancer	Soft Tissue Sarcomas	TP53	0.04
Kidney Cancer	Melanoma	CDKN2A	0.04
Lung Cancer	Bladder Cancer	XPC	0.04
Ovarian Cancer	Lymphoma	TP53	0.04
Bowel Cancer	Lymphoma	TP53,MET	0.05
Bowel Cancer	Leukaemia	BUB1B,MYO18B	0.05
Bowel Cancer	Oesophageal Cancer	MLH3,PTGS2	0.05
Lung Cancer	Soft Tissue Sarcomas	TP53	0.06
Lung Cancer	Mouth Oropharyngeal Cancer	XRCC4	0.06
Breast Cancer	Lymphoma	ATM,TP53	0.07
Pancreatic Cancer	Ovarian Cancer	BRCA2,PALB2	0.07
Bowel Cancer	Bladder Cancer	XPC,ERCC6	0.08
Bowel Cancer	Prostate Cancer	BRCA1,BRCA2,CHEK2,EPHB2,PARP1	0.08
Lung Cancer	Lymphoma	ATM,TP53	0.09
Lung Cancer	Head and neck cancer	EGFR,PTPN13	0.09
Lung Cancer	Womb Cancer	MGMT	0.09
Skin Cancer(Non Melanoma)	Melanoma	MC1R	0.10
Ovarian Cancer	Soft Tissue Sarcomas	TP53	0.11
Stomach Cancer	Ovarian Cancer	TP53,CDH1	0.11
Stomach Cancer	Brain tumours	TP53	0.12
Stomach Cancer	Womb Cancer	MLH1	0.12
Bowel Cancer	Stomach Cancer	TP53,MET,NUDT1,MLH1	0.14
Bowel Cancer	Womb Cancer	MSH6,BRCA1,MLH3,MLH1	0.14
Head and neck cancer	Melanoma	P14ARF	0.14
Melanoma	Leukaemia	RMI1	0.14
Lung Cancer	Skin Cancer(Non Melanoma)	TP53,TP53BP1	0.15
Skin Cancer(Non Melanoma)	Lung Cancer	ERCC2	0.15
Head and neck cancer	Pancreatic Cancer	P14ARF	0.16
Skin Cancer(Non Melanoma)	Stomach Cancer	TP53	0.16
Skin Cancer(Non Melanoma)	Brain tumours	TP53	0.16
Brain tumours	Lymphoma	TP53	0.20
Womb Cancer	Oesophageal Cancer	MLH3	0.20
Prostate Cancer	Ovarian Cancer	BRCA1,BRCA2,CHEK2,CDH1	0.21
Skin Cancer(Non Melanoma)	Lymphoma	TP53	0.24
Kidney Cancer	Pancreatic Cancer	VHL,CDKN2A	0.25
Breast Cancer	Prostate Cancer	BRCA1,BRCA2,CHEK2,PARP1,AR,CDH1,CYP19A1	0.27
Kidney Cancer	Brain tumours	MSH2,SDHB	0.29
Breast Cancer	Soft Tissue Sarcomas	TP53	0.31
Soft Tissue Sarcomas	Brain tumours	TP53	0.31
Stomach Cancer	Soft Tissue Sarcomas	TP53	0.31
Breast Cancer	Ovarian Cancer	BARD1,TP53,BRCA1,BRCA2,CHEK2,CDH1,MRE11A,RAD50,RAD51C,RAD51D	0.37
Skin Cancer(Non Melanoma)	Soft Tissue Sarcomas	TP53	0.38
Melanoma	Pancreatic Cancer	BRCA2,P14ARF,CDKN2A	0.39
Stomach Cancer	Lymphoma	TP53,MET	0.46
Soft Tissue Sarcomas	Lymphoma	TP53	0.49

Table A.1.2: Small Deletions SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 19 genes and 12 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Genes	Kappa
Bowel Cancer	Breast Cancer	CDKN2A	-0.51
Bowel Cancer	Ovarian Cancer	MSH2	-0.24
Bowel Cancer	Pancreatic Cancer	BRCA2,BRCA1,PALB2	-0.14
Breast Cancer	Pancreatic Cancer	MSH2,MSH6,BRCA1,MLH1	-0.08
Lung Cancer	Breast Cancer	MSH2	-0.04
Lung Cancer	Ovarian Cancer	BRCA2	-0.04
Prostate Cancer	Ovarian Cancer	CDKN2A	-0.04
Womb Cancer	Ovarian Cancer	MSH2,MSH6	-0.04
Lung Cancer	Bowel Cancer	MSH2,MSH6	0.10
Breast Cancer	Stomach Cancer	BRCA2,SERCA2B	0.11
Lung Cancer	Pancreatic Cancer	BRCA2,BRCA1,PALB2	0.18
Prostate Cancer	Pancreatic Cancer	MSH2	0.18
Pancreatic Cancer	Ovarian Cancer	BRCA2,SERCA2B	0.19
Breast Cancer	Prostate Cancer	MSH2,MSH6	0.22
Bowel Cancer	Lymphoma	BRCA2,BRCA1,BRIP1,CDH1,CHEK2,RAD50,RAD51C,TP53	0.24
Bowel Cancer	Brain tumours	BRCA2,SERCA2B	0.24
Bowel Cancer	Bladder Cancer	BRCA2	0.24
Melanoma	Pancreatic Cancer	BRCA2,BRCA1,BRIP1,CDH1,CHEK2,RAD50,RAD51C,TP53	0.32
Breast Cancer	Ovarian Cancer	BRCA2,BRCA1,BRIP1,CDH1,CHEK2,RAD50,RAD51C,TP53	0.42
Lung Cancer	Prostate Cancer	CDKN2A	0.42
Bowel Cancer	Womb Cancer	MSH2,MSH6,BRCA1,MLH1	0.46
Womb Cancer	Lymphoma	BRCA2,BRCA1,PALB2	0.63
Womb Cancer	Brain tumours	BRCA2	0.63
Lymphoma	Brain tumours	MSH2,MSH6,BRCA1,MLH1	1.00

Table A.1.3: Splicing SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 12 genes and 12 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Genes	Kappa
Prostate Cancer	Ovarian Cancer	BRCA2	-0.09
Breast Cancer	Pancreatic Cancer	BRCA2	0.00
Pancreatic Cancer	Ovarian Cancer	BRCA2	0.06
Breast Cancer	Prostate Cancer	BRCA2,CHEK2	0.17
Bowel Cancer	Stomach Cancer	MUTYH	0.25
Prostate Cancer	Pancreatic Cancer	BRCA2	0.25
Stomach Cancer	Prostate Cancer	CDH1	0.25
Brain tumours	Pancreatic Cancer	CDKN2A	0.40
Bowel Cancer	Womb Cancer	MSH6	0.43
Bowel Cancer	Kidney Cancer	MSH2	0.43
Bowel Cancer	Bladder Cancer	MSH2	0.43
Melanoma	Brain tumours	CDKN2A	0.63
Melanoma	Pancreatic Cancer	CDKN2A	0.63
Skin Cancer(Non Melanoma)	Brain tumours	SUFU	0.63
Breast Cancer	Ovarian Cancer	BRCA1,BRCA2,RAD51C,RAD51D,TP53	0.83
Kidney Cancer	Bladder Cancer	MSH2	1.00

Table A.1.4: Small Insertion SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 9 genes and 10 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Genes	Kappa
Bowel Cancer	Breast Cancer	CDH1	-0.35
Breast Cancer	Stomach Cancer	CDH1	0.05
Melanoma	Ovarian Cancer	BRCA2	0.05
Pancreatic Cancer	Ovarian Cancer	BRCA2	0.05
Bowel Cancer	Brain tumours	TP53	0.27
Bowel Cancer	Kidney Cancer	MSH2	0.27
Bowel Cancer	Womb Cancer	MSH6	0.27
Lung Cancer	Ovarian Cancer	BRCA2	0.27
Prostate Cancer	Ovarian Cancer	BRCA2	0.27
Bowel Cancer	Stomach Cancer	CDH1,MSH2	0.53
Breast Cancer	Ovarian Cancer	BRCA1,BRIP1,RAD51C	0.55
Lung Cancer	Melanoma	BRCA2	0.61
Lung Cancer	Pancreatic Cancer	BRCA2	0.61
Melanoma	Prostate Cancer	BRCA2	0.61
Prostate Cancer	Pancreatic Cancer	BRCA2	0.61
Stomach Cancer	Kidney Cancer	MSH2	0.61
Lung Cancer	Prostate Cancer	BRCA2	1
Melanoma	Pancreatic Cancer	BRCA2,CDKN2A	1

Table A.1.5: Regulatory SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 14 genes and 11 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Gene	Kappa
Breast Cancer	Bowel Cancer	DNMT3B, ERBB4	-0.29
Bladder Cancer	Breast Cancer	MDM2	-0.14
Bowel Cancer	Bladder Cancer	PTGS2	-0.09
Breast Cancer	Prostate Cancer	KLK3	-0.08
Lung Cancer	Bladder Cancer	PTGS2	-0.02
Breast Cancer	Ovarian Cancer	BRCA2	0.00
Lung Cancer	Breast Cancer	DNMT3B, MIR30C1	0.00
Lung Cancer	Ovarian Cancer	CHEK2	0.10
Skin Cancer (Non Melanoma)	Breast Cancer	TP53	0.14
Bowel Cancer	Oesophageal Cancer	PTGS2	0.19
Bowel Cancer	Womb Cancer	MSH2	0.19
Prostate Cancer	Bowel Cancer	MYC	0.22
Lung Cancer	Oesophageal Cancer	PTGS2	0.24
Lung Cancer	Bowel Cancer	DNMT3B, MLH1, PTGS2	0.26
Bladder Cancer	Oesophageal Cancer	PTGS2	0.44
Bladder Cancer	Stomach Cancer	OGG1	0.44
Cervical Cancer	Prostate Cancer	BCL2	0.44

Table A.1.6: Gross Deletions SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 8 genes and 8 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Genes	Kappa
Bowel Cancer	Ovarian Cancer	BRCA1	-0.07
Breast Cancer	Prostate Cancer	CHEK2	0.00
Prostate Cancer	Ovarian Cancer	BRCA2	0.14
Bowel Cancer	Breast Cancer	BRCA1, TP73	0.25
Breast Cancer	Ovarian Cancer	BRCA1, TP53	0.25
Bowel Cancer	Womb Cancer	MSH2	0.38
Bowel Cancer	Bladder Cancer	MSH2	0.38
Melanoma	Brain tumours	CDKN2A	1.00
Womb Cancer	Bladder Cancer	MSH2	1.00

Table A.1.7: Small Indels SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 5 genes and 8 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Gene	Kappa
Bowel Cancer	Brain tumours	MLH1	0.55
Bowel Cancer	Womb Cancer	MSH6	0.55
Breast Cancer	Ovarian Cancer	BRCA1	0.55
Pancreatic Cancer	Ovarian Cancer	BRCA2	0.55
Head and neck cancer	Melanoma	CDKN2A	1

Table A.1.8: Repeated Variations SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 2 genes and 3 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Gene	Kappa
Breast Cancer	Prostate Cancer	CYP11A1	0
Lung Cancer	Prostate Cancer	TERT	0

Table A.1.9: Complex Rearrangement SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 1 genes and 2 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Shared Genes	Kappa
Breast Cancer	Ovarian Cancer	BRCA1	0

Table A.1.10: Gross Insertions SM-HCM. The first column represents cancer node one, the second column is the associated cancer node for cancers in column one, the third column is the shared gene for each of the two associated cancer nodes, the fourth column is the kappa agreement test results where low agreements <0.25, moderate agreement 0.25 to 0.50, high agreement 0.50 to 0.75 and the very high agreement 0.75 – 1. In total there are 1 gene and 2 primary cancer nodes.

Cancer Node 1	Cancer Node 2	Shared Genes	Kappa
Breast Cancer	Ovarian Cancer	BRCA1	0

Appendix 2

Table A.2.11: Chromosome gene mutations and associated cancers. The first column represents the chromosome number, the second column shows the genes symbols for gene on this particular chromosome, the third column contains the mutation class for genes on the chromosome in column one, and the fourth column shows is the cancers associated with that particular chromosome.

Chromosomes	Genes	Mutations	Cancers
Ch 5	MIR146A,AMACR,GHR,DROSHA,CDH12,ROPN1L,MTRR,TERT,MSH3,XRCC4,PIK3R1,VCAN,APC,LOX,RAD50,IL12B	Regulatory,Missense nonsense,Splicing,Repeat Variations,Small Deletion,Gross Deletions	Prostate cancer,Lung Cancer,Ovarian Cancer,Breast Cancer,Pancreatic Cancer,Leukaemia,Bowel Cancer,Bladder Cancer,Mouth Oropharyngeal Cancer,Stomach Cancer
Ch 19	MIR125A,JAK3,B3GNT3,GDF15,MIR27A,TYK2,ICAM5,STK11,INSR,ICAM1,CEBPA,RHPN2,XRCC1,ERCC1,ERCC2,CD3EAP,GLTSCR1,ZNF350,ZNF600,KLK3,ZNF577	Regulatory,Missense nonsense,Splicing,Small Deletion,Small Insertions	Leukaemia,Lymphoma,Prostate cancer,Stomach Cancer,Bowel Cancer,Ovarian Cancer,Lung Cancer,Skin Cancer(Non Melanoma),Brain tumours,Bladder Cancer
Ch 10	KLF6,AKR1C3,RET,MSMB,ERCC6,PRF1,SFTPD,BMPR1A,ANXA11,PTEN,HIF1AN,FAS,CYP17A1,SUFU,GSTO2,GPAM,DMBT1,MGMT,FGFR2,C10ORF137,PAK7,DNMT3B,GHRH,MIR499,NCOA3,AURKA	Missense nonsense,Splicing,Small Insertions,Small Indels,Comples Rearrangements,Regulatory,Small Deletion,Gross Deletions,Repeat Variations	Lung Cancer,Thyroid Cancer,Bowel Cancer,Bladder Cancer,Lymphoma,Breast Cancer,Head and neck cancer,Leukaemia,Skin Cancer(Non Melanoma),Brain tumours,Ovarian Cancer,Womb Cancer,Oesophageal Cancer
Ch 11	ACCS,SPI1,PTPRJ,WT1,RAG1,NAV2,CSNK2A1P,HRAS,MUC2,MUC6,INS,GTF2H1,FEN1,MS4A64,YAP1,GSTP1,TBX10,MYO7A,MRE11A,MMP7,ATM,PGR,NPAT,CASP5,IL10RA,PPP2R1B,HSPA8	Missense nonsense,Regulatory,Small Insertions,Small Indels,Splicing,Small Deletion,Gross Deletions,Repeat Variations	Leukaemia,Thyroid Cancer,Wilms Tumour,Lymphoma,Bowel Cancer,Bladder Cancer,Stomach Cancer,Prostate cancer,Mouth Oropharyngeal Cancer,Melanoma,Breast Cancer,Ovarian Cancer,Womb Cancer
Ch 12	KRAS,CDKN1B,WNK1,DCP1B,KRT5,ATF1,ITGA7,VDR,MIR196A2,CDK4,MDM2,IGF1,SERCA2B,MYBPC1,PTPN11,SETD8,CHFR	Regulatory,Missense nonsense,Splicing,Repeat Variations,Small Deletion,Gross Insertions	Melanoma,Prostate cancer,Bowel Cancer,Ovarian Cancer,Skin Cancer(Non Melanoma),Bladder Cancer,Breast Cancer,Brain tumours
Ch 13	FLT3,LNX2,BRCA2,FREM2,SETDB2,LPAR6,RB1,EDNRB,MIR17,ERCC5,LIG4	Missense nonsense,Regulatory,Small Deletion,Small Insertions,Splicing,Gross Deletions,Small Indels,Gross Insertions,Comples Rearrangements	Lymphoma,Lung Cancer,Bowel Cancer,Breast Cancer,Melanoma,Prostate cancer,Pancreatic Cancer,Ovarian Cancer,Ocular Cancer,Brain tumours

Ch 14	APEX1,BCL2L2,NFKBIA,NKX2-1,LGALS3,BMP4,MTHFD1,MLH3,GPR68,GOLGA5,DICER1,XRCC3,CDC42BPB	Regulatory,Missense nonsense,Small Deletion,Small Indels,Small Insertions,Splicing	Bowel Cancer,Head and neck cancer,Stomach Cancer,Lymphoma,Thyroid Cancer,Breast Cancer,Melanoma,Womb Cancer,Oesophageal Cancer,Wilms Tumour,Soft Tissue Sarcomas,Brain tumours,Ovarian Cancer
Ch 15	SPRED1,RYR3,BUB1B,TP53BP1,RAD51,CYP19A1,PML,CYP1A1,ITGA11,CYP11A1,CHRNA3,AKAP13,BCL2A1,IQGAP1	Missense nonsense,Splicing,Small Deletion,Small Insertions,Regulatory,Repeat Variations	Breast Cancer,Bowel Cancer,Leukaemia,Lung Cancer,Skin Cancer(Non Melanoma),Prostate cancer
Ch 16	SULT1A1,PALB2,SLX4,UBE2I,CREBBP,ERCC4,LOC643714,TOX3,MMP2,NOD2,NQO1,CDH1,ZFHX3,C16ORF61,WWOX,MC1R,FANCA	Missense nonsense,Splicing,Small Deletion,Gross Deletions,Regulatory,Small Insertions,Small Indels	Pancreatic Cancer,Ovarian Cancer,Breast Cancer,Lung Cancer,Bladder Cancer,Bowel Cancer,Stomach Cancer,Prostate cancer,Lymphoma,Skin Cancer(Non Melanoma),Melanoma
Ch 17	AKAP10,ZNF624,FLCN,MAP2K4,ELAC2,DNAH9,ITGAE,GEMIN4,SHBG,GUCY2D,PFAS,TP53,PLD2,WRAP53,MYH8,CAMKK1,RPH3AL,ATP2A3,RAD51D,NF1,NOS2,BRCA1,PHB,ACACA,ERBB2,CDC6,MPP3,CDK5RAP3,RAD51C,BRIP1,AXIN2,EPX,SSTR2,GH1,BIRC5,FASN	Missense nonsense,Splicing,Regulatory,Small Deletion,Small Insertions,Gross Deletions,Repeat Variations,Small Indels,Gross Insertions,Comples Rearrangements	Bowel Cancer,Kidney Cancer,Prostate cancer,Ovarian Cancer,Breast Cancer,Skin Cancer(Non Melanoma),Stomach Cancer,Soft Tissue Sarcomas,Myxoma,Brain tumours,Lymphoma,Head and neck cancer,Leukaemia,Womb Cancer,Pancreatic Cancer
Ch 18	MEP1B,SMAD7,SMAD4,DCC,BCL2	Missense nonsense,Regulatory	Breast Cancer,Cervical Cancer,Prostate cancer
Ch 1	GSTM1,GSTM3,DDX20,MSH4,IL23R,RAD54L,JUN,PTCH2,MIR30C1,MYCL1,MUTYH,PLA2G2A,RUNX3,EPHB2,SDHB,LIN28A,CA6,MIIP,TP73,CHD1L,ARHGEF11,MUC1,S100A14,DUSP23,MPZ,RNASEL,CDC73,PTGS2,CHIT1,MDM4,ESRRG,TGFB2,PARP1,EXO1,FH	Regulatory,Gross Deletions,Missense nonsense,Splicing,Small Deletion,Small Indels,Small Insertions,Repeat Variations,Gross Insertions	Stomach Cancer,Head and neck cancer,Bladder Cancer,Lung Cancer,Bowel Cancer,Skin Cancer(Non Melanoma),Kidney Cancer,Prostate cancer,Brain tumours,Ovarian Cancer,Oesophageal Cancer,Soft Tissue Sarcomas,Parathyroid carcinoma
Ch 21	NRIP1,GRIK1,IFNGR2,RUNX1,RRP1B,COL18A1	Missense nonsense,Small Deletion	Head and neck cancer,Leukaemia,Breast Cancer,Prostate cancer
Ch 22	SMARCB1,MYO18B,COMT,MIF,CABIN1,CHEK2,NF2,PDGFB,SUN2,TNFRSF13C,SMC1B	Missense nonsense,Gross Deletions,Small Insertions,Regulatory,Repeat Variations,Splicing,Small Deletion,Small Indels,Gross Insertions,Comples Rearrangements	Kidney Cancer,Bowel Cancer,Leukaemia,Breast Cancer,Stomach Cancer,Soft Tissue Sarcomas,Lung Cancer,Prostate cancer,Ovarian Cancer,Lymphoma,Head and neck cancer
Ch 2	MSH6,CYP1B1,EPCAM,LHCGR,FSHR,MSH2,DNMT3A,SRD5A2,IL1R1,IL1B,ERCC3,IFIH1,LRP2,RNASEL,PMS1,PDE11A,CTLA4,CASP8,CFLAR,IGFBP5,ERBB4,BARD1,XRCC5,SLC11A1	Missense nonsense,Splicing,Small Deletion,Regulatory,Small Insertions,Small Indels,Gross Deletions,Gross Insertions,Comples Rearrangements,Repeat Variations	Breast Cancer,Womb Cancer,Ovarian Cancer,Lung Cancer,Prostate cancer,Stomach Cancer,Kidney Cancer,Bladder Cancer,Lymphoma,Brain tumours,Head and neck cancer,Cervical Cancer,Oesophageal Cancer

Ch 3	DHX36,ARL6IP5,IL17RB,GPX1,MLH1,RASSF1,MIR191,KLHDC8B,BAP1,TGFBR2,PLCD1,RAF1,XPC,PPARG,FANCD2,VHL,OGG1,NR1I2,CD86,GTF2E1,DIRC2,GATA2,TOPBP1,IL12A,TNFSF10,SERPINI2,SST,MFI2	Missense nonsense,Regulatory,Small Insertions,Splicing,Small Deletion,Small Indels,Gross Deletions,Gross Insertions,Comples Rearrangements,Repeat Variations	Bladder Cancer,Lung Cancer,Brain tumours,Stomach Cancer,Womb Cancer,Ovarian Cancer,Lymphoma,Melanoma,Leukaemia,Head and neck cancer,Breast Cancer,Oesophageal Cancer,Kidney Cancer,Pancreatic Cancer
Ch 4	TLR1,TLR10,TLR6,BST1,KCNIP4,RAB28,FGFRL1,KIT,PPAT,IGFBP7,UGT2B7,IL8,SULT1E1,TMPRSS11A,PTPN13,ABCG2,BMPR1B,ADH7,AGXT2L1,EGF,CCNA2,MAD2L1,EDNRA,TLR2,PDGFC,PALLD,CASP3	Missense nonsense,Comples Rearrangements,Regulatory,Small Deletion,Gross Deletions,Splicing	Prostate cancer,Lymphoma,Kidney Cancer,Breast Cancer,Lung Cancer,Leukaemia,Head and neck cancer,Bladder Cancer,Stomach Cancer,Oesophageal Cancer,Melanoma,Ovarian Cancer,Pancreatic Cancer
Ch 6	VEGFA,POLH,TRERF1,CDKN1A,BAG6,DHX16,HLA-DQA1,PSMB9,MICA,CCHCR1,C6ORF15,PRRC2A,FANCE,DTNBP1,NQO2,COL12A1,SNORD50A,ASCC3,HACE1,LIN28B,REV3L,ESR1,SOD2	Regulatory,Missense nonsense,Small Indels,Repeat Variations,Small Insertions,Small Deletion,Comples Rearrangements,Splicing	Melanoma,Bowel Cancer,Breast Cancer,Lung Cancer,Mouth Oropharyngeal Cancer,Lymphoma,Oesophageal Cancer,Wilms Tumour,Ovarian Cancer
Ch 7	EGFR,IGFBP3,GHRHR,BBS9,AOAH,POU6F2,MAD1L1,NUDT1,PMS2,CYP2W1,POR,HSPB1,PMS2P3,AKAP9,ABCB1,CYP3A5,HGF,CYP3A4,PON1,SHFM1,MET,CPA4,EPHB6,PRKAG2,XRCC2	Missense nonsense,Repeat Variations,Regulatory,Comples Rearrangements,Gross Deletions,Splicing,Small Deletion,Small Insertions,Small Indels,Gross Insertions	Breast Cancer,Head and neck cancer,Bowel Cancer,Wilms Tumour,Stomach Cancer,Pancreatic Cancer,Prostate cancer,Lymphoma
Ch 8	TNFRSF10A,NKX3-1,MTUS1,LZTS1,MSR1,TNFRSF10B,NAT1,NBN,POP1,CSMD3,RNF139,TG,PVT1,EIF2C2,EIF3H,EXT1,PSCA,POU5F1B,MYC,PYCRL	Missense nonsense,Regulatory,Gross Deletions,Splicing,Small Insertions,Small Indels,Comples Rearrangements,Small Deletion	Bladder Cancer,Prostate cancer,Ovarian Cancer,Breast Cancer,Head and neck cancer,Bowel Cancer,Melanoma,Kidney Cancer,Thyroid Cancer,Bone Cancer,Stomach Cancer
Ch 9	CDKN2A,P14ARF,CDKN2B,CDKN2AI3,SMARCA2,CER1,RMI1,TGFBR1,HSD17B3,PTCH1,PTCH1AM,GALNT12,RAD23B,DEC1,TNC,TNFSF8,EHMT1	Small Indels,Missense nonsense,Splicing,Regulatory,Small Deletion,Small Insertions,Gross Deletions,Gross Insertions,Repeat Variations	Kidney Cancer,Melanoma,Brain tumours,Pancreatic Cancer,Lung Cancer,Leukaemia,Bowel Cancer,Breast Cancer,Prostate cancer,Skin Cancer(Non Melanoma)
Ch X	KDM6A,FAM123B,AR,SH2D1A,FMR1	Missense nonsense,Regulatory,Repeat Variations,Gross Deletions	Wilms Tumour,Breast Cancer,Prostate cancer,Lymphoma,Ovarian Cancer
Ch 20	DNMT3B,GHRH,PAK7,AURKA,DNMT3B,MIR499,NCOA3	Splicing,Regulatory,Missense_nonsense, Repeat Variations	Lung_Cancer, Bowel_Cancer,Breast_Cancer,Oesophageal_Cancer

Appendix 3

Table A.3.1: A sample table of 2,851 collected Missense/Nonsense mutation records, where the first column represents the genes , the second column the chromosome location, the third column the cancer disorder, the fourth column the encoded amino acid (the letters before the hyphen indicates the original codon sequence and the letters after the hyphen the incorrect codon sequence), the fifth column shows the encoded amino acid (the first three letters is the abbreviation of the original amino acid and the second three letters is the abbreviation for the iniquitous encoded amino acid; the numbers in the middle represent the peptide position of the changed amino acid, and the sixth column provides the PubMed literatures for each of the mutation record.

Genes	Locations	Cancer	Codon Change	Amino acid changes	Reference
ATM	11q22-q23	Breast_Cancer	gAGC-CGC	Ser788Arg	Atencio (2001) Environ Mol Mutagen 38, 200
ATM	11q22-q23	Breast_Cancer	tCAT-TAT	His1380Tyr	Atencio (2001) Environ Mol Mutagen 38, 200
ATM	11q22-q23	Breast_Cancer	GAC-GGC	Asp1467Gly	Atencio (2001) Environ Mol Mutagen 38, 200
ATM	11q22-q23	Breast_Cancer	cATT-GTT	Ile2030Val	Atencio (2001) Environ Mol Mutagen 38, 200
ATM	11q22-q23	Breast_Cancer	GCA-GTA	Ala2466Val	Atencio (2001) Environ Mol Mutagen 38, 200
BRCA 1	17q21	Breast_Cancer	aCAA-TAA	Gln12Term	Adem (2003) Cancer 97, 1
BRCA 1	17q21	Breast _&_Ovarian_Can cer	gTGT-CGT	Cys61Arg	Al-Mulla (2009) J Clin Pathol 62, 350
BRCA 1	17q21	Ovarian_cancer	ACA-AGA	Thr276Arg	Akbari (2011) J Med Genet 48, 783
BRCA 1	17q21	Ovarian_cancer	tGAT-AAT	Asp330Asn	Akbari (2011) J Med Genet 48, 783
BRCA 1	17q21	Breast_Cancer	AGGc-AGC	Arg331Ser	Asadi (2008) Scand J Clin Lab Invest 68, 563
BRCA 1	17q21	Breast _&_Ovarian_Can cer	TCA-TGA	Ser603Term	Al-Mulla (2009) J Clin Pathol 62, 350
BRCA 1	17q21	Breast _&_Ovarian_Can cer	cAAG-CAG	Lys918Gln	Al-Mulla (2009) J Clin Pathol 62, 350
BRCA 1	17q21	Breast_Cancer	tGGA-TGA	Gly972Term	Ahn (2007) Cancer Lett 245, 90
BRCA 1	17q21	Breast_Cancer	TTG-TAG	Leu1198Ter m	Ahn (2007) Cancer Lett 245, 90
BRCA 1	17q21	Ovarian_cancer	CAT-CTT	His1421Leu	Akbari (2011) J Med Genet 48, 783

BRCA 1	17q21	Ovarian_cancer	AGG-ATG	Arg1495Met	Aziz (2001) Gynecol Oncol 80, 341
BRCA 1	17q21	Ovarian_cancer	GAT-GTT	Asp1692Val	Akbari (2011) J Med Genet 48, 783
BRCA 1	17q21	Ovarian_cancer	aGGT-AGT	Gly1748Ser	Aktas (2010) Gynecol Oncol epub, epub
BRCA 2	13q12.3	Breast _&_Ovarian_Cancer	aGAT-TAT	Asp23Tyr	Al-Mulla (2009) J Clin Pathol 62, 350
BRCA 2	13q12.3	Breast_Cancer	CGT-CAT	Arg2034His	Awadelkarim (2007) Breast Cancer Res Treat 102, 189
BRCA 2	13q12.3	Ovarian_cancer	aCTT-TTT	Leu2686Phe	Akbari (2011) J Med Genet 48, 783
BRCA 2	13q12.3	Breast_Cancer	TATt-TAG	Tyr2997Term	Ahn (2007) Cancer Lett 245, 90
BRCA 2	13q12.3	Ovarian_cancer	GAT-GGT	Asp3142Gly	Akbari (2011) J Med Genet 48, 783
CDH1	16q22.1	Gastric_cancer	tGCA-ACA	Ala617Thr	Ascano (2001) Mod Pathol 14, 942
RB1	13q14.2	Retinoblastoma	AGG-AAG	Arg46Lys	Abouzeid (2009) Mol Vis 15, 771
RB1	13q14.2	Retinoblastoma	aGAA-TAA	Glu280Term	Babenko (2002) Mol Biol (Mosk) 36, 623
RB1	13q14.2	Retinoblastoma	CTT-CAT	Leu303His	Babenko (2002) Mol Biol (Mosk) 36, 623
RB1	13q14.2	Retinoblastoma	AAA-AGA	Lys319Arg	Babenko (2002) Mol Biol (Mosk) 36, 623
RB1	13q14.2	Retinoblastoma	aCAG-TAG	Gln436Term	Abouzeid (2009) Mol Vis 15, 771
RB1	13q14.2	Retinoblastoma	TGCg-TGA	Cys489Term	Babenko (2002) Mol Biol (Mosk) 36, 623
RB1	13q14.2	Retinoblastoma	TATa-TAA	Tyr498Term	Babenko (2002) Mol Biol (Mosk) 36, 623
RB1	13q14.2	Retinoblastoma	gAAA-TAA	Lys615Term	Abouzeid (2009) Mol Vis 15, 771
RB1	13q14.2	Retinoblastoma	aGAA-TAA	Glu746Term	Babenko (2002) Mol Biol (Mosk) 36, 623
RB1	13q14.2	Retinoblastoma	TTA-TAA	Leu797Term	Ata-ur-Rasheed (2002) Ophthalmic Genet 23, 121
RB1	13q14.2	Retinoblastoma	TTCc-TTA	Phe845Leu	Babenko (2002) Mol Biol (Mosk) 36, 623
RET	10q11.2	Thyroid_cancer	cAAG-GAG	Lys666Glu	Ahmed (2005) J Mol Diagn 7, 283
MLH1	3p21.3	Colorectal_cancer	GTG-GAG	Val49Glu	Auclair (2006) Hum Mutat 27, 145
MLH1	3p21.3	Colorectal_cancer	AGCa-AGA	Ser106Arg	Auclair (2006) Hum Mutat 27, 145
MLH1	3p21.3	Colorectal_cancer	GTT-GAT	Val113Asp	Auclair (2006) Hum Mutat 27, 145
MLH1	3p21.3	Colorectal_cancer	gGGG-CGG	Gly454Arg	Auclair (2006) Hum Mutat 27, 145
MLH1	3p21.3	Colorectal_cancer	GCC-GAC	Ala586Asp	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	cACA-GCA	Thr33Ala	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	TAT-TGT	Tyr43Cys	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	CTT-CGT	Leu187Arg	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	CGG-CAG	Arg243Gln	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	TCT-TGT	Ser323Cys	Akiyama (1997) Biochem Biophys Res Commun 236, 248
MSH2	2p22-p21	Colorectal_cancer	CTT-CCT	Leu341Pro	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	tGGT-TGT	Gly548Cys	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	tGAA-AAA	Glu561Lys	Auclair (2006) Hum Mutat 27, 145
MSH2	2p22-p21	Colorectal_cancer	AAAg-AAC	Lys627Asn	Auclair (2006) Hum Mutat 27, 145
TP53	17p13.1	Breast_&_Colorectal_Cancer	GGC-GAC	Gly244Asp	Achatz (2007) Cancer Lett 245, 96
TP53	17p13.1	Choroid_Plexus_Tumours	AAC-ATC	Asn131Ile	Agarwalla (2008) Pediatr Neurosurg 44, 501
TP53	17p13.1	Colorectal_cancer	aGTG-ATG	Val197Met	Achatz (2007) Cancer Lett 245, 96

Table A.3.2: The 21 altered amino acid network tables, where the first column is node one gene, the second column is the associated node two gene, the third column is the cancer disorder name, the fourth column is a match column where '0' is equal to identical altered amino acid in both genes, '1' is the identical codon change in both genes and '2' is the identical altered amino acid occurs in same peptide position of the two altered amino acid. Fifth and sixth columns are the encoded codon and amino acid for node gene one, and the seventh and eight columns are the encoded codon and amino acid for node gene two.

Alanine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	0	GTT-GCT GTT-GCT GGA-GCA GTA-GCA	Val1570Ala Val1729Ala Gly2287Ala Val2439Ala	aCCC-GCC tCCT-GCT	Pro151Ala Pro278Ala
BRCA1	TP53	Breast Cancer	1	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala	aCCC-GCC tCCT-GCT	Pro151Ala Pro278Ala
BRCA1	BRIP1	Breast Cancer	1	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala	tCCC-GCC	Pro47Ala
BRCA1	ATM	Breast Cancer	1	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala	GTT-GCT GTT-GCT GGA-GCA GTA-GCA	Val1570Ala Val1729Ala Gly2287Ala Val2439Ala
BRCA1	PALB2	Breast Cancer	1	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala	GGT-GCT	Gly1043Ala

				GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala		
BRCA2	ATM	Breast Cancer	1	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala	GTT-GCT GTT-GCT GGA-GCA GTA-GCA	Val1570Ala Val1729Ala Gly2287Ala Val2439Ala
BRCA2	BRCA1	Breast Cancer	1	GTC-GCC tACT-GCT aACA-GCA GAT-GCT tACT-GCT aCCC-GCC GAT-GCT GAG-GCG tACC-GCC	Val211Ala Thr225Ala Thr598Ala Asp1990Ala Thr2031Ala Pro2283Ala Asp2723Ala Glu2856Ala Thr3349Ala	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala
BRCA2	PALB2	Breast Cancer	0	GTC-GCC tACT-GCT aACA-GCA GAT-GCT tACT-GCT aCCC-GCC GAT-GCT GAG-GCG tACC-GCC	Val211Ala Thr225Ala Thr598Ala Asp1990Ala Thr2031Ala Pro2283Ala Asp2723Ala Glu2856Ala Thr3349Ala	GGT-GCT	Gly1043Ala
BRCA2	TP53	Breast Cancer	1	GTC-GCC tACT-GCT aACA-GCA GAT-GCT tACT-GCT aCCC-GCC GAT-GCT GAG-GCG tACC-GCC	Val211Ala Thr225Ala Thr598Ala Asp1990Ala Thr2031Ala Pro2283Ala Asp2723Ala Glu2856Ala Thr3349Ala	aCCC-GCC tCCT-GCT	Pro151Ala Pro278Ala
CHEK2	BRCA2	Breast Cancer	1	tACA-GCA GGG-GCG	Thr172Ala Gly306Ala	GTC-GCC tACT-GCT aACA-GCA GAT-GCT tACT-GCT aCCC-GCC GAT-GCT GAG-GCG tACC-GCC	Val211Ala Thr225Ala Thr598Ala Asp1990Ala Thr2031Ala Pro2283Ala Asp2723Ala Glu2856Ala Thr3349Ala
CHEK2	ATM	Breast Cancer	0	tACA-GCA GGG-GCG	Thr172Ala Gly306Ala	GTT-GCT GTT-GCT GGA-GCA GTA-GCA	Val1570Ala Val1729Ala Gly2287Ala Val2439Ala
CHEK2	TP53	Breast Cancer	0	tACA-GCA GGG-GCG	Thr172Ala Gly306Ala	aCCC-GCC tCCT-GCT	Pro151Ala Pro278Ala
CHEK2	BRCA1	Breast Cancer	1	tACA-GCA GGG-GCG	Thr172Ala Gly306Ala	GTA-GCA gCCT-GCT aCCC-GCC tTCT-GCT GGT-GCT aACA-GCA aACG-GCG GTA-GCA GTA-GCA GCAt-GCG aCCA-GCA tACT-GCT aACA-GCA gACA-GCA GGA-GCA GTA-GCA GTC-GCC tCCA-GCA GTT-GCT gCCA-GCA	Val11Ala Pro58Ala Pro142Ala Ser186Ala Gly263Ala Thr557Ala Thr582Ala Val772Ala Val1047Ala Ala1142Ala Pro1491Ala Thr1685Ala Thr1691Ala Thr1700Ala Gly1706Ala Val1713Ala Val1736Ala Pro1749Ala Val1809Ala Pro1812Ala
ERCC5	ERCC3	Lung Cancer	0	aACG-GCG	Thr971Ala	cACT-GCT	Thr232Ala
EXO1	MLH1	Bowel Cancer	0	GTA-GCA cCCC-GCC	Val27Ala Pro640Ala	GGG-GCG tACT-GCT aCCA-GCA GTG-GCG aACA-GCA GTT-GCT GAT-GCT	Gly22Ala Thr82Ala Pro141Ala Val326Ala Thr364Ala Val506Ala Asp631Ala

						GAG-GCG	Glu663Ala
EXO1	MSH2	Bowel Cancer	0	GTA-GCA cCCC-GCC	Val27Ala Pro640Ala	cACA-GCA GGA-GCA tACC-GCC GGT-GCT GAG-GCG	Thr33Ala Gly162Ala Thr564Ala Gly674Ala Glu853Ala
EXO1	MSH3	Bowel Cancer	0	GTA-GCA cCCC-GCC	Val27Ala Pro640Ala	cACA-GCA	Thr1045Ala
MSH2	MLH1	Bowel Cancer	1	cACA-GCA GGA-GCA tACC-GCC GGT-GCT GAG-GCG	Thr33Ala Gly162Ala Thr564Ala Gly674Ala Glu853Ala	GGG-GCG tACT-GCT aCCA-GCA GTG-GCG aACA-GCA GTT-GCT GAT-GCT GAG-GCG	Gly22Ala Thr82Ala Pro141Ala Val326Ala Thr364Ala Val506Ala Asp631Ala Glu663Ala
MSH3	MLH1	Bowel Cancer	0	cACA-GCA	Thr1045Ala	GGG-GCG tACT-GCT aCCA-GCA GTG-GCG aACA-GCA GTT-GCT GAT-GCT GAG-GCG	Gly22Ala Thr82Ala Pro141Ala Val326Ala Thr364Ala Val506Ala Asp631Ala Glu663Ala
MSH3	MSH2	Bowel Cancer	1	cACA-GCA	Thr1045Ala	cACA-GCA GGA-GCA tACC-GCC GGT-GCT GAG-GCG	Thr33Ala Gly162Ala Thr564Ala Gly674Ala Glu853Ala
MSH6	MLH1	Bowel Cancer	1	GGG-GCG gCCA-GCA GTG-GCG GGT-GCT GTT-GCT	Gly54Ala Pro202Ala Val509Ala Gly685Ala Val878Ala	GGG-GCG tACT-GCT aCCA-GCA GTG-GCG aACA-GCA GTT-GCT GAT-GCT GAG-GCG	Gly22Ala Thr82Ala Pro141Ala Val326Ala Thr364Ala Val506Ala Asp631Ala Glu663Ala
MSH6	MSH2	Bowel Cancer	1	GGG-GCG gCCA-GCA GTG-GCG GGT-GCT	Gly54Ala Pro202Ala Val509Ala Gly685Ala	cACA-GCA GGA-GCA tACC-GCC GGT-GCT GAG-GCG	Thr33Ala Gly162Ala Thr564Ala Gly674Ala Glu853Ala
Arginine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	0	gAGC-CGC CAG-CGG CAT-CGT aTGT-CGT aGGA-AGA tGGA-AGA CCT-CGT	Ser788Arg Gln1128Arg His2111Arg Cys2464Arg Gly2772Arg Gly2023Arg Pro1054Arg	cTGC-CGC	Cys242Arg
BRCA1	PALB2	Breast Cancer	1	gTGT-CGT ACA-AGA gTGT-CGT CAC-CGC tTGC-CGC AAA-AGA gTGT-CGT aTGT-CGT AGCc-AGA ACG-AGG aGGT-CGT AGTt-AGG CAG-CGG tGGG-AGG gTGT-CGT CAG-CGG AGCa-AGG AAA-AGA aGGG-CGG CCT-CGT CAC-CGC ATG-AGG ACA-AGA AGCt-AGA tAGC-CGC aGGA-AGA CCA-CGA ATG-AGG CAG-CGG gTGG-CGG	Cys24Arg Thr37Arg Cys39Arg His41Arg Cys44Arg Lys56Arg Cys61Arg Cys64Arg Ser72Arg Thr77Arg Gly98Arg Ser264Arg Gln356Arg Gly462Arg Cys805Arg Gln867Arg Ser1024Arg Lys1183Arg Gly1205Arg Pro1238Arg His1284Arg Met1689Arg Thr1691Arg Ser1715Arg Ser1715Arg Gly1738Arg Pro1749Arg Met1775Arg Gln1811Arg Trp1837Arg	CCA-CGA AAG-AGG	Pro207Arg Lys1048Arg
BRCA2	BUB1B	Leukaemia	1	CAA-CGA	Gln1396Arg	CAA-CGA	Gln349Arg
BRCA2	BRCA1	Breast Cancer	1	AAA-AGA CAT-CGT CCA-CGA AGCa-AGA CAG-CGG ACA-AGA AAG-AGG CTT-CGT cGGG-AGG AAA-AGA AGTa-AGA tGGT-CGT aTGG-CGG CTT-CGT ACA-AGA	Lys53Arg His150Arg Pro201Arg Ser326Arg Gln499Arg Thr1011Arg Lys1057Arg Leu1114Arg Gly1529Arg Lys1530Arg Ser2006Arg Gly2353Arg Trp2626Arg Leu2721Arg Thr2722Arg	gTGT-CGT ACA-AGA gTGT-CGT CAC-CGC tTGC-CGC AAA-AGA gTGT-CGT aTGT-CGT AGCc-AGA ACG-AGG aGGT-CGT AGTt-AGG CAG-CGG tGGG-AGG gTGT-CGT CAG-CGG AGCa-AGG AAA-AGA aGGG-CGG CCT-CGT CAC-CGC ATG-AGG ACA-AGA gTGT-CGT AGCt-AGA tAGC-CGC aGGA-AGA CCA-CGA	Cys24Arg Thr37Arg Cys39Arg His41Arg Cys44Arg Lys56Arg Cys61Arg Cys64Arg Ser72Arg Thr77Arg Gly98Arg Ser264Arg Gln356Arg Gly462Arg Cys805Arg Gln867Arg Ser1024Arg Lys1183Arg Gly1205Arg Pro1238Arg His1284Arg Met1689Arg Thr1691Arg Cys1697Arg Ser1715Arg Ser1715Arg Gly1738Arg Pro1749Arg

						ATG-AGG CAG-CGG gTGG-CGG	Met1775Arg Gln1811Arg Trp1837Arg
BRCA2	BRCA1	Ovarian Cancer	1	CAG-CGG cGGG-AGG AAA-AGA AGTa-AGA CAC-CGC tGGT-CGT	Gln499Arg Gly1529Arg Lys1530Arg Ser2006Arg His2116Arg Gly2353Arg	ATG-AGG ACA-AGA gTGT-CGT CAC-CGC AAA-AGA gTGT-CGT AGCc-AGA ACA-AGA CAG-CGG tGGG-AGG CCT-CGT ATG-AGG ACA-AGA AGCt-AGA tAGC-CGC aGGA-AGA CAG-CGG	Met1Arg Thr37Arg Cys39Arg His41Arg Lys56Arg Cys61Arg Ser72Arg Thr276Arg Gln356Arg Gly462Arg Pro1238Arg Met1689Arg Thr1691Arg Ser1715Arg Ser1715Arg Gly1738Arg Gln1811Arg
BRCA2	PALB2	Breast Cancer	1	AAA-AGA CAT-CGT CCA-CGA AGCa-AGA CAG-CGG ACA-AGA AAG-AGG CTT-CGT cGGG-AGG AAA-AGA AGTa-AGA CAC-CGC tGGT-CGT aTGG-CGG CTT-CGT ACA-AGA	Lys53Arg His150Arg Pro201Arg Ser326Arg Gln499Arg Thr1011Arg Lys1057Arg Leu1114Arg Gly1529Arg Lys1530Arg Ser2006Arg His2116Arg Gly2353Arg Trp2626Arg Leu2721Arg Thr2722Arg	CCA-CGA AAG-AGG	Pro207Arg Lys1048Arg
CHEK2	BRCA2	Breast Cancer	0	CCT-CGT	Pro85Arg	AAA-AGA CCA-CGA AGCa-AGA ACA-AGA AAG-AGG CTT-CGT aTGG-CGG CTT-CGT ACA-AGA	Lys53Arg Pro201Arg Ser326Arg Thr1011Arg Lys1057Arg Leu1114Arg Trp2626Arg Leu2721Arg Thr2722Arg
CHEK2	ATM	Breast Cancer	1	CCT-CGT	Pro85Arg	gAGC-CGC CAG-CGG CAT-CGT aTGT-CGT aGGA-AGA tGGA-AGA CCT-CGT	Ser788Arg Gln1128Arg His2111Arg Cys2464Arg Gly2772Arg Gly2023Arg Pro1054Arg
CHEK2	TP53	Breast Cancer	0	CCT-CGT	Pro85Arg	cTGC-CGC	Cys242Arg
CHEK2	BRCA1	Breast Cancer	1	CCT-CGT	Pro85Arg	gTGT-CGT ACA-AGA gTGT-CGT CAC-CGC tTGC-CGC AAA-AGA gTGT-CGT aTGT-CGT AGCc-AGA ACG-AGG aGGT-CGT AGTt-AGG CAG-CGG tGGG-AGG gTGT-CGT CAG-CGG AGCa-AGG AAA-AGA aGGG-CGG CCT-CGT CAC-CGC ATG-AGG ACA-AGA gTGT-CGT aCGG-AGG AGCt-AGA tAGC-CGC aGGA-AGA CCA-CGA ATG-AGG CAG-CGG gTGG-CGG	Cys24Arg Thr37Arg Cys39Arg His41Arg Cys44Arg Lys56Arg Cys61Arg Cys64Arg Ser72Arg Thr77Arg Gly98Arg Ser264Arg Gln356Arg Gly462Arg Cys805Arg Gln867Arg Ser1024Arg Lys1183Arg Gly1205Arg Pro1238Arg His1284Arg Met1689Arg Thr1691Arg Cys1697Arg Arg1699Arg Ser1715Arg Ser1715Arg Gly1738Arg Pro1749Arg Met1775Arg Gln1811Arg Trp1837Arg
CHEK2	BRCA1	Ovarian Cancer	1	CAC-CGC	His143Arg	ATG-AGG ACA-AGA gTGT-CGT CAC-CGC AAA-AGA gTGT-CGT AGCc-AGA ACA-AGA CAG-CGG tGGG-AGG CCT-CGT ATG-AGG ACA-AGA AGCt-AGA tAGC-CGC aGGA-AGA CAG-CGG	Met1Arg Thr37Arg Cys39Arg His41Arg Lys56Arg Cys61Arg Ser72Arg Thr276Arg Gln356Arg Gly462Arg Pro1238Arg Met1689Arg Thr1691Arg Ser1715Arg Ser1715Arg Gly1738Arg Gln1811Arg
ERCC2	ERCC3	Lung Cancer	0	CGCt-CGA	Arg156Arg	AAG-AGG	Lys449Arg
EXO1	MLH1	Bowel Cancer	1	CTC-CGC	Leu410Arg	ATG-AGG ATG-AGG cGGG-AGG cGGG-CGG aTGT-CGT tGGC-CGC AGCa-AGA ATA-AGA ACG-AGG aGGG-CGG	Met1Arg Met35Arg Gly67Arg Gly67Arg Cys77Arg Gly98Arg Ser106Arg Ile107Arg Thr117Arg Gly147Arg

						CTT-CGT CTC-CGC CAT-CGT gGGG-CGG CTG-CGG CTT-CGT CCC-CGC aGGG-AGG CTG-CGG tTGG-CGG CTC-CGC aTGC-CGC CAG-CGG AAA-AGA	Leu155Arg Leu260Arg His264Arg Gly454Arg Leu559Arg Leu585Arg Pro603Arg Gly634Arg Leu653Arg Trp666Arg Leu676Arg Cys680Arg Gln689Arg Lys751Arg
EXO1	PMS2	Bowel Cancer	0	CTC-CGC	Leu410Arg	CCT-CGT	Pro404Arg
EXO1	MSH2	Bowel Cancer	0	CTC-CGC	Leu410Arg	tAGA-CGA tGGA-AGA tGGA-CGA tGGG-AGG CTT-CGT CTT-CGT gTGT-CGT aGGA-AGA CCT-CGT cAGC-CGC aGGC-CGC CAA-CGA CAT-CGT tGGC-CGC aGGT-CGT tGGG-AGG ATG-AGG tGGG-CGG aTGT-CGT gGGA-AGA CAT-CGT ACA-AGA CTA-CGA	Arg99Arg Gly162Arg Gly162Arg Gly164Arg Leu187Arg Leu310Arg Cys333Arg Gly338Arg Pro349Arg Ser554Arg Gly587Arg Gln629Arg His639Arg Gly669Arg Gly674Arg Gly683Arg Met688Arg Gly692Arg Cys697Arg Gly751Arg His839Arg Thr905Arg Leu911Arg
MLH3	MLH1	Bowel Cancer	1	aTGG-CGG	Trp1276Arg	ATG-AGG ATG-AGG cGGG-AGG cGGG-CGG aTGT-CGT tGGC-CGC AGCa-AGA ATA-AGA ACG-AGG aGGG-CGG CTT-CGT CTC-CGC CAT-CGT gGGG-CGG CTG-CGG CTT-CGT CCC-CGC aGGG-AGG CTG-CGG tTGG-CGG CTC-CGC aTGC-CGC CAG-CGG AAA-AGA	Met1Arg Met35Arg Gly67Arg Gly67Arg Cys77Arg Gly98Arg Ser106Arg Ile107Arg Thr117Arg Gly147Arg Leu155Arg Leu260Arg His264Arg Gly454Arg Leu559Arg Leu585Arg Pro603Arg Gly634Arg Leu653Arg Trp666Arg Leu676Arg Cys680Arg Gln689Arg Lys751Arg
MSH2	MLH1	Gastric Cancer	0	CTA-CGA	Leu173Arg	AGCa-AGG	Ser106Arg
MSH2	MLH1	Bowel Cancer	2	tAGA-CGA tGGA-AGA tGGA-CGA tGGG-AGG CTT-CGT CTT-CGT gTGT-CGT aGGA-AGA CCT-CGT cAGC-CGC aGGC-CGC CAA-CGA CAT-CGT tGGC-CGC aGGT-CGT tGGG-AGG ATG-AGG tGGG-CGG aTGT-CGT gGGA-AGA CAT-CGT ACA-AGA CTA-CGA	Arg99Arg Gly162Arg Gly162Arg Gly164Arg Leu187Arg Leu310Arg Cys333Arg Gly338Arg Pro349Arg Ser554Arg Gly587Arg Gln629Arg His639Arg Gly669Arg Gly674Arg Gly683Arg Met688Arg Gly692Arg Cys697Arg Gly751Arg His839Arg Thr905Arg Leu911Arg	ATG-AGG ATG-AGG cGGG-AGG cGGG-CGG aTGT-CGT tGGC-CGC AGCa-AGA ATA-AGA ACG-AGG aGGG-CGG CTT-CGT CTC-CGC CAT-CGT gGGG-CGG CTG-CGG CTT-CGT CCC-CGC aGGG-AGG CTG-CGG tTGG-CGG CTC-CGC aTGC-CGC CAG-CGG AAA-AGA	Met1Arg Met35Arg Gly67Arg Gly67Arg Cys77Arg Gly98Arg Ser106Arg Ile107Arg Thr117Arg Gly147Arg Leu155Arg Leu260Arg His264Arg Gly454Arg Leu559Arg Leu585Arg Pro603Arg Gly634Arg Leu653Arg Trp666Arg Leu676Arg Cys680Arg Gln689Arg Lys751Arg
MSH6	MLH1	Bowel Cancer	1	AAA-AGA CAG-CGG gGGA-AGA tGGG-AGG	Lys295Arg Gln522Arg Gly566Arg Gly670Arg	ATG-AGG ATG-AGG cGGG-AGG cGGG-CGG aTGT-CGT tGGC-CGC AGCa-AGA ATA-AGA ACG-AGG aGGG-CGG CTT-CGT CTC-CGC CAT-CGT gGGG-CGG CTG-CGG CTT-CGT CCC-CGC aGGG-AGG CTG-CGG tTGG-CGG CTC-CGC aTGC-CGC CAG-CGG AAA-AGA	Met1Arg Met35Arg Gly67Arg Gly67Arg Cys77Arg Gly98Arg Ser106Arg Ile107Arg Thr117Arg Gly147Arg Leu155Arg Leu260Arg His264Arg Gly454Arg Leu559Arg Leu585Arg Pro603Arg Gly634Arg Leu653Arg Trp666Arg Leu676Arg Cys680Arg Gln689Arg Lys751Arg
MSH6	MSH2	Bowel Cancer	1	AAA-AGA CAG-CGG	Lys295Arg Gln522Arg	tAGA-CGA tGGA-AGA	Arg99Arg Gly162Arg

				gGGA-AGA tGGG-AGG	Gly566Arg Gly670Arg	tGGA-CGA tGGG-AGG CTT-CGT CTT-CGT gTGT-CGT aGGA-AGA CCT-CGT cAGC-CGC aGGC-CGC CAA-CGA CAT-CGT tGGC-CGC aGGT-CGT tGGG-AGG ATG-AGG tGGG-CGG aTGT-CGT gGGA-AGA CAT-CGT ACA-AGA CTA-CGA	Gly162Arg Gly164Arg Leu187Arg Leu310Arg Cys333Arg Gly338Arg Pro349Arg Ser554Arg Gly587Arg Gln629Arg His639Arg Gly669Arg Gly674Arg Gly683Arg Met688Arg Gly692Arg Cys697Arg Gly751Arg His839Arg Thr905Arg Leu911Arg
MSH6	MSH2	Bowel Cancer	1	AAA-AGA CAG-CGG gGGA-AGA tGGG-AGG	Lys295Arg Gln522Arg Gly566Arg Gly670Arg	tAGA-CGA tGGA-AGA tGGA-CGA tGGG-AGG CTT-CGT CTT-CGT gTGT-CGT aGGA-AGA CCT-CGT cAGC-CGC aGGC-CGC CAA-CGA CAT-CGT tGGC-CGC aGGT-CGT tGGG-AGG ATG-AGG tGGG-CGG aTGT-CGT gGGA-AGA CAT-CGT ACA-AGA CTA-CGA	Arg99Arg Gly162Arg Gly162Arg Gly164Arg Leu187Arg Leu310Arg Cys333Arg Gly338Arg Pro349Arg Ser554Arg Gly587Arg Gln629Arg His639Arg Gly669Arg Gly674Arg Gly683Arg Met688Arg Gly692Arg Cys697Arg Gly751Arg His839Arg Thr905Arg Leu911Arg
MUTYH	MSH6	Bowel Cancer	1	gTGG-AGG CAG-CGG	Trp131Arg Gln338Arg	CAG-CGG gGGA-AGA tGGG-AGG	Gln522Arg Gly566Arg Gly670Arg
PARP1	TP53	Breast Cancer	0	cCGA-AGA	Arg452Arg	cTGC-CGC	Cys242Arg
PMS1	MLH1	Bowel Cancer	0	gGGA-AGA	Gly501Arg	ATG-AGG ATG-AGG cGGG-AGG cGGG-CGG aTGT-CGT tGGC-CGC AGCa-AGA ATA-AGA ACG-AGG aGGG-CGG CTT-CGT CTC-CGC CAT-CGT gGGG-CGG CTG-CGG CTT-CGT CCC-CGC aGGG-AGG CTG-CGG tTGG-CGG CTC-CGC aTGC-CGC CAG-CGG AAA-AGA	Met1Arg Met35Arg Gly67Arg Gly67Arg Cys77Arg Gly98Arg Ser106Arg Ile107Arg Thr117Arg Gly147Arg Leu155Arg Leu260Arg His264Arg Gly454Arg Leu559Arg Leu585Arg Pro603Arg Gly634Arg Leu653Arg Trp666Arg Leu676Arg Cys680Arg Gln689Arg Lys751Arg
PMS2	MSH2	Bowel Cancer	1	CCT-CGT	Pro404Arg	tAGA-CGA tGGA-AGA tGGA-CGA tGGG-AGG CTT-CGT CTT-CGT gTGT-CGT aGGA-AGA CCT-CGT cAGC-CGC aGGC-CGC CAA-CGA CAT-CGT tGGC-CGC aGGT-CGT tGGG-AGG ATG-AGG tGGG-CGG aTGT-CGT gGGA-AGA CAT-CGT ACA-AGA CTA-CGA	Arg99Arg Gly162Arg Gly162Arg Gly164Arg Leu187Arg Leu310Arg Cys333Arg Gly338Arg Pro349Arg Ser554Arg Gly587Arg Gln629Arg His639Arg Gly669Arg Gly674Arg Gly683Arg Met688Arg Gly692Arg Cys697Arg Gly751Arg His839Arg Thr905Arg Leu911Arg
PMS2	MLH1	Bowel Cancer	0	CCT-CGT	Pro404Arg	ATG-AGG ATG-AGG cGGG-AGG cGGG-CGG aTGT-CGT tGGC-CGC AGCa-AGA ATA-AGA ACG-AGG aGGG-CGG CTT-CGT CTC-CGC CAT-CGT gGGG-CGG CTG-CGG CTT-CGT CCC-CGC aGGG-AGG CTG-CGG	Met1Arg Met35Arg Gly67Arg Gly67Arg Cys77Arg Gly98Arg Ser106Arg Ile107Arg Thr117Arg Gly147Arg Leu155Arg Leu260Arg His264Arg Gly454Arg Leu559Arg Leu585Arg Pro603Arg Gly634Arg Leu653Arg

						tTGG-CGG CTC-CGC aTGC-CGC CAG-CGG AAA-AGA	Trp666Arg Leu676Arg Cys680Arg Gln689Arg Lys751Arg
PTEN	TP53	Breast Cancer	0	AGTg-AGG	Ser207Arg	cTGC-CGC	Cys242Arg
Asparagine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
BRCA1	TP53	Breast Cancer	1	AAGt-AAT AAAt-AAT AAAg-AAT tGAC-AAC AGT-AAT cGAT-AAT AGC-AAC AGT-AAT AGC-AAC cCAC-AAC AGT-AAT	Lys38Asn Lys45Asn Lys56Asn Asp96Asn Ser264Asn Asp695Asn Ser1040Asn Ser1389Asn Ser1715Asn His1746Asn Ser1841Asn	aGAC-AAC	Asp281Asn
BRCA1	BARD1	Breast Cancer	1	AAGt-AAT AAAt-AAT AAAg-AAT tGAC-AAC AGT-AAT cGAT-AAT AGC-AAC AGT-AAT AGC-AAC cCAC-AAC AGT-AAT	Lys38Asn Lys45Asn Lys56Asn Asp96Asn Ser264Asn Asp695Asn Ser1040Asn Ser1389Asn Ser1715Asn His1746Asn Ser1841Asn	AAAt-AAT	Lys312Asn
BRCA1	ATM	Breast Cancer	1	AAGt-AAT AAAt-AAT AAAg-AAT tGAC-AAC AGT-AAT cGAT-AAT AGC-AAC AGT-AAT AGC-AAC cCAC-AAC AGT-AAT	Lys38Asn Lys45Asn Lys56Asn Asp96Asn Ser264Asn Asp695Asn Ser1040Asn Ser1389Asn Ser1715Asn His1746Asn Ser1841Asn	AGT-AAT gGAC-AAC AAAa-AAC aGAT-AAT	Ser759Asn Asn1099Asn Lys1454Asn Asp1853Asn
BRCA2	BARD1	Breast Cancer	0	tGAT-AAT tGAT-AAT AGC-AAC AAAa-AAC aCAC-AAC AAAg-AAC ACT-AAT AAGg-AAT	Asp244Asn Asp935Asn Ser1179Asn Lys1690Asn His2074Asn Lys2128Asn Thr2310Asn Lys2950Asn	AAAt-AAT	Lys312Asn
BRCA2	ATM	Breast Cancer	1	AAGt-AAT AAAt-AAT AAAg-AAT tGAC-AAC AGT-AAT cGAT-AAT AGC-AAC AGT-AAT AGC-AAC cCAC-AAC AGT-AAT	Lys38Asn Lys45Asn Lys56Asn Asp96Asn Ser264Asn Asp695Asn Ser1040Asn Ser1389Asn Ser1715Asn His1746Asn Ser1841Asn	AGT-AAT gGAC-AAC AAAa-AAC aGAT-AAT	Ser759Asn Asn1099Asn Lys1454Asn Asp1853Asn
BRCA2	BRCA1	Breast Cancer	1	tGAT-AAT tGAT-AAT AGC-AAC AAAa-AAC aCAC-AAC AAAg-AAC ACT-AAT AAGg-AAT	Asp244Asn Asp935Asn Ser1179Asn Lys1690Asn His2074Asn Lys2128Asn Thr2310Asn Lys2950Asn	AAGt-AAT AAAt-AAT AAAg-AAT tGAC-AAC AGT-AAT cGAT-AAT AGC-AAC AGT-AAT AGC-AAC cCAC-AAC AGT-AAT	Lys38Asn Lys45Asn Lys56Asn Asp96Asn Ser264Asn Asp695Asn Ser1040Asn Ser1389Asn Ser1715Asn His1746Asn Ser1841Asn
BRCA2	BRCA1	Ovarian Cancer	1	tGAT-AAT AAAa-AAC aCAC-AAC AAGg-AAT	Asp935Asn Lys1690Asn His2074Asn Lys2950Asn	AAGt-AAT AAAt-AAT tGAC-AAC tGAT-AAT AGC-AAC cCAC-AAC AGT-AAT	Lys38Asn Lys45Asn Asp96Asn Asn330Asn Ser1715Asn His1746Asn Ser1841Asn
BRCA2	TP53	Breast Cancer	0	tGAT-AAT tGAT-AAT AGC-AAC AAAa-AAC aCAC-AAC AAAg-AAC ACT-AAT AAGg-AAT	Asp244Asn Asp935Asn Ser1179Asn Lys1690Asn His2074Asn Lys2128Asn Thr2310Asn Lys2950Asn	aGAC-AAC	Asp281Asn
CDKN2A	CDK4	Melanoma	0	gCAC-AAC cGAC-AAC cGAT-AAT	His83Asn Asp84Asn Asp108Asn	AGC-AAC	Ser52Asn
MSH2	MLH1	Bowel Cancer	1	aGAT-AAT AAAg-AAC ATC-AAC	Asp603Asn Lys627Asn Ile708Asn	ATT-AAT aGAC-AAC ATC-AAC aTAC-AAC gCAC-AAC AGT-AAT	Ile36Asn Asp63Asn Ile68Asn Tyr126Asn His329Asn Ser406Asn
MSH6	MLH1	Bowel Cancer	1	AAGa-AAT ATT-AAT ACT-AAT AGT-AAT	Lys99Asn Ile745Asn Thr764Asn Ser1188Asn	ATT-AAT aGAC-AAC ATC-AAC aTAC-AAC gCAC-AAC AGT-AAT	Ile36Asn Asp63Asn Ile68Asn Tyr126Asn His329Asn Ser406Asn
MSH6	MSH2	Bowel Cancer	0	AAGa-AAT ATT-AAT ACT-AAT AGT-AAT	Lys99Asn Ile745Asn Thr764Asn Ser1188Asn	aGAT-AAT AAAg-AAC ATC-AAC	Asp603Asn Lys627Asn Ile708Asn
PMS2	MSH2	Bowel Cancer	0	AGT-AAT	Ser46Asn	aGAT-AAT AAAg-AAC ATC-AAC	Asp603Asn Lys627Asn Ile708Asn
PMS2	MLH1	Bowel Cancer	1	AGT-AAT	Ser46Asn	ATT-AAT aGAC-AAC	Ile36Asn Asp63Asn

						ATC-AAC aTAC-AAC gCAC-AAC AGT-AAT	Ile68Asn Tyr126Asn His329Asn Ser406Asn
PTEN	TP53	Breast Cancer	1	aGAC-AAC aGAT-AAT aGAT-AAT aGAT-AAT	Asp92Asn Asp107Asn Asp115Asn Asp153Asn	aGAC-AAC	Asp281Asn
TP53	BARD1	Breast Cancer	0	aGAC-AAC	Asp281Asn	AAAt-AAT	Lys312Asn
				Glutamate			
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
BRCA1	ATM	Breast Cancer	1	GATa-GAG aAAG-GAG cAAA-GAA gAAA-GAA GTG-GAG GGA-GAA GCG-GAG GGA-GAA GATg-GAG GTG-GAG	Asp67Glu Lys654Glu Lys820Glu Lys1606Glu Val1696Glu Gly1706Glu Ala1708Glu Gly1738Glu Asp1739Glu Val1838Glu	GACg-GAA tAAG-GAG tCAA-GAA tAAA-GAA	Asp203Glu Lys1964Glu Gln2277Glu Lys2431Glu
BRCA1	PALB2	Breast Cancer	1	GATa-GAG aAAG-GAG cAAA-GAA gAAA-GAA GTG-GAG GGA-GAA GCG-GAG GGA-GAA GATg-GAG GTG-GAG	Asp67Glu Lys654Glu Lys820Glu Lys1606Glu Val1696Glu Gly1706Glu Ala1708Glu Gly1738Glu Asp1739Glu Val1838Glu	GGA-GAA	Gly998Glu
BRCA2	ATM	Breast Cancer	1	GATa-GAG aAAG-GAG cAAA-GAA gAAA-GAA GTG-GAG GGA-GAA GCG-GAG GGA-GAA GATg-GAG GTG-GAG	Asp67Glu Lys654Glu Lys820Glu Lys1606Glu Val1696Glu Gly1706Glu Ala1708Glu Gly1738Glu Asp1739Glu Val1838Glu	GACg-GAA tAAG-GAG tCAA-GAA tAAA-GAA	Asp203Glu Lys1964Glu Gln2277Glu Lys2431Glu
BRCA2	PALB2	Breast Cancer	0	cAAG-GAG tCAG-GAG GACg-GAG tAAA-GAA	Lys327Glu Gln2456Glu Asp3095Glu Lys3196Glu	GGA-GAA	Gly998Glu
CHEK2	BRCA2	Breast Cancer	1	aAAA-GAA	Lys224Glu	GACg-GAG tAAA-GAA	Asp3095Glu Lys3196Glu
CHEK2	ATM	Breast Cancer	1	aAAA-GAA	Lys224Glu	GACg-GAA tAAG-GAG tCAA-GAA tAAA-GAA	Asp203Glu Lys1964Glu Gln2277Glu Lys2431Glu
CHEK2	BRCA1	Breast Cancer	1	aAAA-GAA	Lys224Glu	GATa-GAG aAAG-GAG cAAA-GAA gAAA-GAA GTG-GAG GGA-GAA GCG-GAG GGA-GAA GATg-GAG GTG-GAG	Asp67Glu Lys654Glu Lys820Glu Lys1606Glu Val1696Glu Gly1706Glu Ala1708Glu Gly1738Glu Asp1739Glu Val1838Glu
EXO1	MLH1	Bowel Cancer	1	GGA-GAA	Gly759Glu	GCG-GAG GTG-GAG GGA-GAA GACa-GAA GGG-GAG tAAA-GAA GTG-GAG tAAG-GAG GATt-GAA gAAG-GAG	Ala21Glu Val49Glu Gly54Glu Asp63Glu Gly67Glu Lys84Glu Val303Glu Lys416Glu Asp485Glu Lys618Glu
EXO1	PMS2	Bowel Cancer	1	GGA-GAA	Gly759Glu	GGA-GAA	Gly207Glu
MLH3	MLH1	Bowel Cancer	1	cCAA-GAA gAAA-GAA	Gln24Glu Lys412Glu	GCG-GAG GTG-GAG GGA-GAA GACa-GAA GGG-GAG tAAA-GAA GTG-GAG tAAG-GAG GATt-GAA gAAG-GAG	Ala21Glu Val49Glu Gly54Glu Asp63Glu Gly67Glu Lys84Glu Val303Glu Lys416Glu Asp485Glu Lys618Glu
MSH2	MLH1	Bowel Cancer	1	GTA-GAA cCAA-GAA GGA-GAA tAAA-GAA GTA-GAA	Val470Glu Gln690Glu Gly759Glu Lys845Glu Val923Glu	GCG-GAG GTG-GAG GGA-GAA GACa-GAA GGG-GAG tAAA-GAA GTG-GAG tAAG-GAG GATt-GAA gAAG-GAG	Ala21Glu Val49Glu Gly54Glu Asp63Glu Gly67Glu Lys84Glu Val303Glu Lys416Glu Asp485Glu Lys618Glu
MSH6	MLH1	Bowel Cancer	1	tCAG-GAG GGG-GAG	Gln698Glu Gly1069Glu	GCG-GAG GTG-GAG GGA-GAA GACa-GAA GGG-GAG tAAA-GAA aGAG-CAG GTG-GAG tAAG-GAG GATt-GAA gAAG-GAG GGG-GAG	Ala21Glu Val49Glu Gly54Glu Asp63Glu Gly67Glu Lys84Glu Glu199Gln Val303Glu Lys416Glu Asp485Glu Lys618Glu Gly39Glu
MSH6	MSH2	Bowel Cancer	0	GGG-GAG tCAG-GAG GGG-GAG	Gly39Glu Gln698Glu Gly1069Glu	GTA-GAA cCAA-GAA GGA-GAA tAAA-GAA GTA-GAA	Val470Glu Gln690Glu Gly759Glu Lys845Glu Val923Glu
MUTYH	MSH6	Bowel Cancer	1	GGG-GAG	Gly286Glu	GGG-GAG	Gly39Glu

				GGG-GAG	Gly216Glu	tCAG-GAG GGG-GAG	Gln698Glu Gly1069Glu
PMS2	MSH2	Bowel Cancer	1	GGA-GAA	Gly207Glu	GTA-GAA cCAA-GAA GGA-GAA tAAA-GAA GTA-GAA	Val470Glu Gln690Glu Glv759Glu Lys845Glu Val923Glu
PMS2	MLH1	Bowel Cancer	1	GGA-GAA	Gly207Glu	GCG-GAG GTG-GAG GGA-GAA GACa-GAA GGG-GAG tAAA-GAA GTG-GAG tAAG-GAG GATt-GAA gAAG-GAG	Ala21Glu Val49Glu Gly54Glu Asp63Glu Gly67Glu Lys84Glu Val303Glu Lys416Glu Asp485Glu Lys618Glu
Glycine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	0	tAGT-GGT tAGT-GGT GAC-GGC GAC-GGC GAA-GGA GAT-GGT cAGA-GGA	Ser99Glv Ser378Glv Asn1467Glv Asn1914Glv Glu2570Glv Asn2625Glv Arg2912Glv	GTG-GGG	Val122Gly
BRCA1	TP53	Breast Cancer	1	gTGT-GGT tTGC-GGC tTGC-GGC gTGT-GGT aTGT-GGT aAGG-GGG GGTa-GGC GAA-GGA GAG-GGG GGTa-GGA GAT-GGT aAGA-GGA gCGA-GGA GAT-GGT GTC-GGC GTG-GGG gTGG-GGG	Cys39Gly Cys44Gly Cys47Gly Cys61Gly Cys64Gly Arg71Gly Gly263Gly Glu421Gly Glu515Gly Gly774Gly Asp1344Gly Arg1347Gly Arg1443Gly Asp1739Gly Val1741Gly Val1810Gly Trp1837Gly	GTG-GGG	Val122Gly
BRCA1	MRE11A	Breast Cancer	1	gTGT-GGT tTGC-GGC tTGC-GGC gTGT-GGT aTGT-GGT aAGG-GGG GGTa-GGC GAA-GGA GAG-GGG GGTa-GGA GAT-GGT aAGA-GGA gCGA-GGA GAT-GGT GTC-GGC GTG-GGG gTGG-GGG	Cys39Glv Cys44Glv Cys47Glv Cys61Glv Cys64Glv Arg71Glv Glv263Glv Glu421Glv Glu515Glv Glv774Glv Asn1344Glv Arg1347Glv Arg1443Glv Asn1739Glv Val1741Glv Val1810Glv Trp1837Glv	gAGA-GGA	Arg202Gly
BRCA1	ATM	Breast Cancer	1	gTGT-GGT tTGC-GGC tTGC-GGC gTGT-GGT aTGT-GGT aAGG-GGG GGTa-GGC GAA-GGA GAG-GGG GGTa-GGA GAT-GGT aAGA-GGA gCGA-GGA GAT-GGT GTC-GGC GTG-GGG gTGG-GGG	Cys39Gly Cys44Gly Cys47Gly Cys61Gly Cys64Gly Arg71Gly Gly263Gly Glu421Gly Glu515Gly Gly774Gly Asp1344Gly Arg1347Gly Arg1443Gly Asp1739Gly Val1741Gly Val1810Gly Trp1837Gly	tAGT-GGT tAGT-GGT GAC-GGC GAC-GGC GAA-GGA GAT-GGT cAGA-GGA	Ser99Gly Ser378Gly Asp1467Gly Asp1914Gly Glu2570Gly Asp2625Gly Arg2912Gly
BRCA2	ATM	Breast Cancer	1	gTGT-GGT tTGC-GGC tTGC-GGC gTGT-GGT aTGT-GGT aAGG-GGG GGTa-GGC GAA-GGA GAG-GGG GGTa-GGA GAT-GGT aAGA-GGA gCGA-GGA GAT-GGT GTC-GGC GTG-GGG gTGG-GGG	Cys39Glv Cys44Glv Cys47Glv Cys61Glv Cys64Glv Arg71Glv Glv263Glv Glu421Glv Glu515Glv Glv774Glv Asn1344Glv Arg1347Glv Arg1443Glv Asn1739Glv Val1741Glv Val1810Glv Trp1837Glv	tAGT-GGT tAGT-GGT GAC-GGC GAC-GGC GAA-GGA GAT-GGT cAGA-GGA	Ser99Glv Ser378Glv Asn1467Glv Asn1914Glv Glu2570Glv Asn2625Glv Arg2912Gly
BRCA2	BRCA1	Breast Cancer	1	aTGG-GGG GAA-GGA GAA-GGA GAA-GGA GTG-GGG GCA-GGA GAT-GGT	Trp395Gly Glu462Gly Glu2029Gly Glu2121Gly Val2280Gly Ala2351Gly Asp2811Gly	gTGT-GGT tTGC-GGC tTGC-GGC gTGT-GGT aTGT-GGT aAGG-GGG GGTa-GGC GAA-GGA GAG-GGG GGTa-GGA GAT-GGT aAGA-GGA gCGA-GGA GAT-GGT GTC-GGC GTG-GGG gTGG-GGG	Cys39Gly Cys44Gly Cys47Gly Cys61Gly Cys64Gly Arg71Gly Gly263Gly Glu421Gly Glu515Gly Gly774Gly Asp1344Gly Arg1347Gly Arg1443Gly Asp1739Gly Val1741Gly Val1810Gly Trp1837Gly

BRCA2	BRCA1	Ovarian Cancer	1	GGTa-GGC GAA-GGA GAT-GGT	Gly1552Glv Glu2121Glv Asp3142Gly	gTGT-GGT tTGC-GGC aAGG-GGG GAT-GGT GAT-GGT GTG-GGG gTGG-GGG	Cys39Glv Cys44Glv Arg71Glv Asn1344Glv Asn1739Glv Val1810Glv Trp1837Glv
BRCA2	TP53	Breast Cancer	1	aTGG-GGG GAA-GGA GGTa-GGC GAA-GGA GAA-GGA GTG-GGG GCA-GGA GAT-GGT	Trp395Gly Glu462Gly Gly1552Gly Glu2029Gly Glu2121Gly Val2280Gly Ala2351Gly Asp2811Gly	GTG-GGG	Val122Gly
CHEK2	BRCA2	Breast Cancer	0	gAGG-GGG	Arg117Gly	aTGG-GGG GAA-GGA GAA-GGA GTG-GGG GCA-GGA GAT-GGT	Trp395Glv Glu462Glv Glu2029Glv Val2280Glv Ala2351Glv Asp2811Glv
CHEK2	ATM	Breast Cancer	0	gAGG-GGG	Arg117Gly	tAGT-GGT tAGT-GGT GAC-GGC GAC-GGC GAA-GGA GAT-GGT cAGA-GGA	Ser99Gly Ser378Gly Asp1467Gly Asp1914Gly Glu2570Gly Asp2625Gly Arg2912Glv
CHEK2	TP53	Breast Cancer	0	gAGG-GGG	Arg117Glv	GTG-GGG	Val122Glv
CHEK2	BRCA1	Breast Cancer	0	gAGG-GGG	Arg117Gly	gTGT-GGT tTGC-GGC tTGC-GGC gTGT-GGT aTGT-GGT aAGG-GGG GGTa-GGC GAA-GGA GAG-GGG GAT-GGT aAGA-GGA gCGA-GGA GAT-GGT GTC-GGC GTG-GGG gTGG-GGG	Cys39Gly Cys44Gly Cys47Gly Cys61Gly Cys64Gly Arg71Gly Gly263Gly Glu421Gly Glu515Gly Asp1344Gly Arg1347Gly Arg1443Gly Asp1739Gly Val1741Gly Val1810Gly Trp1837Glv
EXO1	MLH1	Bowel Cancer	1	tAGT-GGT	Ser610Gly	GCT-GGT GAT-GGT cAGT-GGT cAGG-GGG GTA-GGA GAG-GGG GAA-GGA GCC-GGC GAT-GGT GAG-GGG GAT-GGT GAT-GGT gAGG-GGG	Ala29Glv Asp41Glv Ser93Glv Arg182Glv Val185Glv Glu234Glv Glu268Glv Ala282Glv Asp304Glv Glu578Glv Asp591Glv Asp601Glv Arg755Gly
EXO1	MSH2	Bowel Cancer	0	tAGT-GGT	Ser610Gly	GTT-GGT GAA-GGA GAT-GGT tTGT-GGT GAT-GGT GAG-GGG GAG-GGG	Val163Gly Glu198Gly Asp603Gly Cys641Gly Asp660Gly Glu853Gly Glu886Glv
MLH3	MLH1	Bowel Cancer	1	tAGT-GGT	Ser817Gly	GCT-GGT GAT-GGT cAGT-GGT cAGG-GGG GTA-GGA GAG-GGG GAA-GGA GCC-GGC GAT-GGT GAG-GGG GAT-GGT GAT-GGT gAGG-GGG	Ala29Glv Asp41Glv Ser93Glv Arg182Glv Val185Glv Glu234Glv Glu268Glv Ala282Glv Asp304Glv Glu578Glv Asp591Glv Asp601Glv Arg755Glv
MRE11A	ATM	Breast Cancer	1	gAGA-GGA	Arg202Gly	tAGT-GGT tAGT-GGT GAC-GGC GAC-GGC GAA-GGA GAT-GGT cAGA-GGA	Ser99Gly Ser378Gly Asp1467Gly Asp1914Gly Glu2570Gly Asp2625Gly Arg2912Glv
MSH2	MLH1	Bowel Cancer	1	GTT-GGT GAA-GGA GAT-GGT tTGT-GGT GAT-GGT GAG-GGG GAG-GGG	Val163Glv Glu198Glv Asp603Glv Cys641Glv Asp660Glv Glu853Glv Glu886Gly	GCT-GGT GAT-GGT cAGT-GGT cAGG-GGG GTA-GGA GAG-GGG GAA-GGA GCC-GGC GAT-GGT GAG-GGG GAT-GGT GAT-GGT gAGG-GGG	Ala29Glv Asp41Glv Ser93Glv Arg182Glv Val185Glv Glu234Glv Glu268Glv Ala282Glv Asp304Glv Glu578Glv Asp591Glv Asp601Glv Arg755Glv
MSH6	MLH1	Bowel Cancer	0	gAGA-GGA GAC-GGC aAGA-GGA	Arg635Gly Asp803Gly Arg1321Gly	GCT-GGT GAT-GGT cAGT-GGT cAGG-GGG GTA-GGA GAG-GGG GAA-GGA GCC-GGC GAT-GGT GAG-GGG GAT-GGT GAT-GGT gAGG-GGG GAG-GGG	Ala29Gly Asp41Gly Ser93Gly Arg182Gly Val185Gly Glu234Gly Glu268Gly Ala282Gly Asp304Gly Glu578Gly Asp591Gly Asp601Gly Arg755Gly Glu229Glv

MSH6	MSH2	Bowel Cancer	1	GAG-GGG gAGA-GGA GAC-GGC aAGA-GGA	Glu229Gly Arg635Gly Asp803Gly Arg1321Gly	GTT-GGT GAA-GGA GAT-GGT tTGT-GGT GAT-GGT GAG-GGG GAG-GGG	Val163Gly Glu198Gly Asp603Gly Cys641Gly Asp660Gly Glu853Gly Glu886Gly
PTEN	TP53	Bowel Cancer	1	GGTg-GGG GTG-GGG	Gly132Gly Val191Gly	GTG-GGG	Val122Gly
Histidine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	1	aTAT-CAT CTC-CAC CGT-CAT CCT-CAT CGT-CAT CGT-CAT	Tyr1442His Leu1555His Arg1575His Pro1785His Arg2719His Arg3008His	CTC-CAC CGC-CAC	Leu130His Arg181His
BRCA1	TP53	Breast Cancer	0	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His	CTC-CAC CGC-CAC	Leu130His Arg181His
BRCA1	BARD1	Breast Cancer	2	CGT-CAT CGT-CAT gAAT-CAT	Arg496His Arg504His Asn564His	CAGc-CAC	Gln564His
BRCA1	RAD50	Breast Cancer	1	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His	CGT-CAT	Arg224His
BRCA1	ATM	Breast Cancer	2	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His	aTAT-CAT CTC-CAC CGT-CAT CCT-CAT CGT-CAT CGT-CAT	Tyr1442His Leu1555His Arg1575His Pro1785His Arg2719His Arg3008His
BRCA1	PALB2	Breast Cancer	1	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His	aTAT-CAT	Tyr408His
BRCA2	ATM	Breast Cancer	1	CGC-CAC aAAT-CAT aAAT-CAT tGAT-CAT CGT-CAT CGT-CAT tAAC-CAC aGAT-CAT tTAC-CAC	Arg18His Asn289His Asn372His Asp596His Arg2034His Arg2108His Asn2135His Asp2723His Tyr3098His	aTAT-CAT CTC-CAC CGT-CAT CCT-CAT CGT-CAT CGT-CAT	Tyr1442His Leu1555His Arg1575His Pro1785His Arg2719His Arg3008His
BRCA2	BRCA1	Breast Cancer	1	CGC-CAC aAAT-CAT aAAT-CAT tGAT-CAT CGT-CAT CGT-CAT tAAC-CAC aGAT-CAT tTAC-CAC	Arg18His Asn289His Asn372His Asp596His Arg2034His Arg2108His Asn2135His Asp2723His Tyr3098His	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His
BRCA2	TP53	Breast Cancer	1	CGC-CAC aAAT-CAT aAAT-CAT tGAT-CAT CGT-CAT CGT-CAT tAAC-CAC aGAT-CAT tTAC-CAC	Arg18His Asn289His Asn372His Asp596His Arg2034His Arg2108His Asn2135His Asp2723His Tyr3098His	CTC-CAC CGC-CAC	Leu130His Arg181His
CDKN2A	CDK4	Melanoma	0	cGAT-CAT	Asp125His	CGT-CAT	Arg24His
CHEK2	BRCA2	Breast Cancer	1	aTAC-CAC gTAT-CAT	Tyr159His Tyr424His	CGC-CAC aAAT-CAT aAAT-CAT tGAT-CAT CGT-CAT CGT-CAT tAAC-CAC aGAT-CAT tTAC-CAC	Arg18His Asn289His Asn372His Asp596His Arg2034His Arg2108His Asn2135His Asp2723His Tyr3098His
CHEK2	RAD50	Breast Cancer	0	aTAC-CAC gTAT-CAT	Tyr159His Tyr424His	CGT-CAT	Arg224His
CHEK2	ATM	Breast Cancer	0	aTAC-CAC gTAT-CAT	Tyr159His Tyr424His	aTAT-CAT CTC-CAC CGT-CAT CCT-CAT CGT-CAT	Tyr1442His Leu1555His Arg1575His Pro1785His Arg2719His

						CGT-CAT	Arg3008His
CHEK2	TP53	Breast Cancer	0	aTAC-CAC gTAT-CAT	Tyr159His Tyr424His	CTC-CAC CGC-CAC	Leu130His Arg181His
CHEK2	TP53	Breast Cancer	0	CGC-CAC CGT-CAT CGC-CAC	Arg180His Arg181His Arg318His	CAAc-CAT	Gln136His
CHEK2	BRCA1	Breast Cancer	1	aTAC-CAC gTAT-CAT	Tyr159His Tyr424His	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His
MSH2	MLH1	Bowel Cancer	1	gGAT-CAT CGT-CAT CCT-CAT	Asp167His Arg382His Pro652His	gAAC-CAC aGAT-CAT aGAT-CAT CGT-CAT CAGa-CAT tGAT-CAT cTAC-CAC CTT-CAT CTT-CAT CGC-CAC CCT-CAT	Asn38His Asp41His Asp132His Arg265His Gln346His Asp485His Tyr561His Leu607His Leu622His Arg725His Pro747His
MSH6	MLH1	Bowel Cancer	1	CGT-CAT CGT-CAT CGT-CAT CGC-CAC	Arg468His Arg901His Arg976His Arg1095His	gAAC-CAC aGAT-CAT aGAT-CAT CGT-CAT CAGa-CAT tGAT-CAT cTAC-CAC CTT-CAT CTT-CAT CGC-CAC CCT-CAT	Asn38His Asp41His Asp132His Arg265His Gln346His Asp485His Tyr561His Leu607His Leu622His Arg725His Pro747His
MSH6	MSH2	Bowel Cancer	1	CGT-CAT CGT-CAT CGC-CAC	Arg468His Arg976His Arg1095His	gGAT-CAT CGT-CAT CCT-CAT	Asp167His Arg382His Pro652His
MUTYH	APEX1	Bowel Cancer	1	CAGg-CAC	Gln498His	CAGa-CAC	Gln51His
MUTYH	MSH6	Bowel Cancer	0	CAGg-CAC	Gln498His	CGT-CAT CGT-CAT CGT-CAT CGC-CAC	Arg468His Arg901His Arg976His Arg1095His
NCOA3	BRCA1	Breast Cancer	0	CAGt-CAC	Gln586His	CCC-CAC CGT-CAT CGT-CAT tAAT-CAT gAAT-CAT gTAT-CAT CAGg-CAT gTAC-CAC CAAg-CAC aGAT-CAT CAGc-CAT	Pro142His Arg496His Arg504His Asn550His Asn564His Tyr856His Gln1200His Tyr1666His Gln1747His Asp1778His Gln1785His
OGG1	XPC	Bowel Cancer	0	CGC-CAC	Arg154His	CGT-CAT	Arg492His

Isoleucine

Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
ATM	TP53	Breast Cancer	0	ATGg-ATA ATGt-ATA	Met1210Ile Met2482Ile	tGTC-ATC	Val97Ile
BRCA1	TP53	Breast Cancer	0	ATGg-ATT cGTT-ATT ACA-ATA AGA-ATA AAT-ATT tGTT-ATT ATGt-ATA ACA-ATA ACT-ATT AGT-ATT AGT-ATT ACC-ATC ATGg-ATA ACT-ATT	Met1Ile Val191Ile Thr374Ile Arg507Ile Asn909Ile Val990Ile Met1008Ile Thr1025Ile Thr1163Ile Ser1164Ile Ser1512Ile Thr1561Ile Met1652Ile Thr1685Ile	tGTC-ATC	Val97Ile
BRCA1	BRIP1	Breast Cancer	1	ATGg-ATT cGTT-ATT ACA-ATA AGA-ATA AAT-ATT tGTT-ATT ATGt-ATA ACA-ATA ACT-ATT AGT-ATT AGT-ATT ACC-ATC ATGg-ATA ACT-ATT	Met1Ile Val191Ile Thr374Ile Arg507Ile Asn909Ile Val990Ile Met1008Ile Thr1025Ile Thr1163Ile Ser1164Ile Ser1512Ile Thr1561Ile Met1652Ile Thr1685Ile	ATGg-ATA	Met299Ile
BRCA1	ATM	Breast Cancer	0	ATGg-ATT cGTT-ATT ACA-ATA AGA-ATA AAT-ATT tGTT-ATT ATGt-ATA ACA-ATA ACT-ATT AGT-ATT AGT-ATT ACC-ATC ATGg-ATA ACT-ATT	Met1Ile Val191Ile Thr374Ile Arg507Ile Asn909Ile Val990Ile Met1008Ile Thr1025Ile Thr1163Ile Ser1164Ile Ser1512Ile Thr1561Ile Met1652Ile Thr1685Ile	ATGg-ATA ATGt-ATA	Met1210Ile Met2482Ile
BRCA1	PALB2	Breast Cancer	1	ATGg-ATT cGTT-ATT ACA-ATA AGA-ATA AAT-ATT	Met1Ile Val191Ile Thr374Ile Arg507Ile Asn909Ile	ACA-ATA	Thr300Ile

				tGTT-ATT ATGt-ATA ACA-ATA ACT-ATT AGT-ATT AGT-ATT ACC-ATC ATGg-ATA ACT-ATT	Val990Ile Met1008Ile Thr1025Ile Thr1163Ile Ser1164Ile Ser1512Ile Thr1561Ile Met1652Ile Thr1685Ile		
BRCA2	ATM	Breast Cancer	1	ATGg-ATT cGTT-ATT ACA-ATA AGA-ATA AAT-ATT tGTT-ATT ATGt-ATA ACA-ATA ACT-ATT AGT-ATT AGT-ATT ACC-ATC ATGg-ATA ACT-ATT	Met1Ile Val191Ile Thr374Ile Arg507Ile Asn909Ile Val990Ile Met1008Ile Thr1025Ile Thr1163Ile Ser1164Ile Ser1512Ile Thr1561Ile Met1652Ile Thr1685Ile	ATGg-ATA ATGt-ATA	Met1210Ile Met2482Ile
BRCA2	BRCA1	Breast Cancer	2	ATGc-ATA ACT-ATT ACA-ATA AAC-ATC tGTT-ATT ACA-ATA tGTT-ATT ACT-ATT cGTC-ATC AAC-ATC	Met1Ile Thr64Ile Thr200Ile Asn924Ile Val2109Ile Thr2515Ile Val2728Ile Thr3013Ile Val3091Ile Asn3124Ile	ATGg-ATT cGTT-ATT ACA-ATA AGA-ATA AAT-ATT tGTT-ATT ATGt-ATA ACA-ATA ACT-ATT AGT-ATT AGT-ATT ACC-ATC ATGg-ATA ACT-ATT	Met1Ile Val191Ile Thr374Ile Arg507Ile Asn909Ile Val990Ile Met1008Ile Thr1025Ile Thr1163Ile Ser1164Ile Ser1512Ile Thr1561Ile Met1652Ile Thr1685Ile
BRCA2	BRCA1	Ovarian Cancer	1	ACT-ATT ACA-ATA ACA-ATA cGTC-ATC	Thr64Ile Thr200Ile Thr2515Ile Val3091Ile	cGTT-ATT ACA-ATA AGT-ATT ATGg-ATA ACT-ATT	Val191Ile Thr374Ile Ser1164Ile Met1652Ile Thr1685Ile
BRCA2	PALB2	Breast Cancer	1	ATGc-ATA ACT-ATT ACA-ATA AAC-ATC tGTT-ATT ACA-ATA tGTT-ATT ACT-ATT cGTC-ATC AAC-ATC	Met1Ile Thr64Ile Thr200Ile Asn924Ile Val2109Ile Thr2515Ile Val2728Ile Thr3013Ile Val3091Ile Asn3124Ile	ACA-ATA	Thr300Ile
BRCA2	TP53	Breast Cancer	1	ATGc-ATA ACT-ATT ACA-ATA AAC-ATC tGTT-ATT ACA-ATA tGTT-ATT ACT-ATT cGTC-ATC AAC-ATC	Met1Ile Thr64Ile Thr200Ile Asn924Ile Val2109Ile Thr2515Ile Val2728Ile Thr3013Ile Val3091Ile Asn3124Ile	tGTC-ATC	Val97Ile
MET	PIK3R1	Bowel Cancer	0	ACT-ATT	Thr1010Ile	ATGt-ATA	Met326Ile
MLH3	MLH1	Bowel Cancer	0	gGTA-ATA	Val971Ile	ATGt-ATA AGT-ATT ACT-ATT AGT-ATT AGC-ATC aTTT-ATT ACA-ATA AGA-ATA tTTT-ATT	Met1Ile Ser46Ile Thr82Ile Ser191Ile Ser321Ile Phe396Ile Thr413Ile Arg472Ile Phe568Ile
MSH2	MLH1	Bowel Cancer	1	AGC-ATC aGTT-ATT AAT-ATT ACC-ATC tGTT-ATT AAT-ATT ATGg-ATA ATGt-ATT	Ser13Ile Val102Ile Asn127Ile Thr335Ile Val342Ile Asn583Ile Met688Ile Met729Ile	ATGt-ATA AGT-ATT ACT-ATT AGT-ATT AGC-ATC aTTT-ATT ACA-ATA AGA-ATA tTTT-ATT	Met1Ile Ser46Ile Thr82Ile Ser191Ile Ser321Ile Phe396Ile Thr413Ile Arg472Ile Phe568Ile
MSH3	MLH1	Bowel Cancer	0	cGTA-ATA	Val79Ile	ATGt-ATA AGT-ATT ACT-ATT AGT-ATT AGC-ATC aTTT-ATT ACA-ATA AGA-ATA tTTT-ATT	Met1Ile Ser46Ile Thr82Ile Ser191Ile Ser321Ile Phe396Ile Thr413Ile Arg472Ile Phe568Ile
MSH3	MSH2	Bowel Cancer	0	cGTA-ATA	Val79Ile	AGC-ATC aGTT-ATT AAT-ATT ACC-ATC tGTT-ATT AAT-ATT ATGg-ATA ATGt-ATT	Ser13Ile Val102Ile Asn127Ile Thr335Ile Val342Ile Asn583Ile Met688Ile Met729Ile
MSH6	MLH1	Bowel Cancer	1	AGC-ATC AGT-ATT AGT-ATT aTTT-ATT ACT-ATT	Ser144Ile Ser227Ile Ser285Ile Phe596Ile Thr1219Ile	ATGt-ATA AGT-ATT ACT-ATT AGT-ATT AGC-ATC aTTT-ATT ACA-ATA AGA-ATA tTTT-ATT	Met1Ile Ser46Ile Thr82Ile Ser191Ile Ser321Ile Phe396Ile Thr413Ile Arg472Ile Phe568Ile
MSH6	MSH2	Bowel Cancer	1	AGC-ATC AGT-ATT AGT-ATT aTTT-ATT ACT-ATT	Ser144Ile Ser227Ile Ser285Ile Phe596Ile Thr1219Ile	AGC-ATC aGTT-ATT AAT-ATT ACC-ATC tGTT-ATT AAT-ATT	Ser13Ile Val102Ile Asn127Ile Thr335Ile Val342Ile Asn583Ile

						ATGg-ATA ATGt-ATT	Met688Ile Met729Ile
PMS2	MSH2	Bowel Cancer	1	ATGa-ATA	Met622Ile	AGC-ATC aGTT-ATT AAT-ATT ACC-ATC tGTT-ATT AAT-ATT ATGg-ATA ATGt-ATT	Ser13Ile Val102Ile Asn127Ile Thr335Ile Val342Ile Asn583Ile Met688Ile Met729Ile
PMS2	MLH1	Bowel Cancer	1	ATGa-ATA	Met622Ile	ATGt-ATA AGT-ATT ACT-ATT AGT-ATT AGC-ATC aTTT-ATT ACA-ATA AGA-ATA tTTT-ATT	Met1Ile Ser46Ile Thr82Ile Ser191Ile Ser321Ile Phe396Ile Thr413Ile Arg472Ile Phe568Ile
PMS2	MSH3	Bowel Cancer	0	ATGa-ATA	Met622Ile	cGTA-ATA	Val79Ile
TP53	PML	Breast Cancer	0	tGTC-ATC	Val97Ile	ACC-ATC	Thr28Ile
XRCC2	RAD51C	Breast Cancer	0	cCTA-ATA	Leu61Ile	ACC-ATC ACC-ATC	Thr102Ile Thr102Ile
				Lysine			
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	0	tCAG-AAG AGA-AAA AACa-AAG	Gln654Lys Arg1437Lys Asn2435Lys	aGAG-AAG	Glu285Lys
BRCA1	TP53	Breast Cancer	1	ACA-AAA ATA-AAA AGG-AAG tGAA-AAA AACc-AAA cGAA-AAA ACA-AAA AATa-AAG cGAG-AAG aGAG-AAG ACA-AAA tGAA-AAA AGA-AAA ATG-AAG	Thr37Lys Ile68Lys Arg71Lys Glu116Lys Asn132Lys Glu575Lys Thr826Lys Asn1236Lys Glu1250Lys Glu1358Lys Thr1691Lys Glu1735Lys Arg1753Lys Met1775Lys	aGAG-AAG	Glu285Lys
BRCA1	ATM	Breast Cancer	1	ACA-AAA ATA-AAA AGG-AAG tGAA-AAA AACc-AAA cGAA-AAA ACA-AAA AATa-AAG cGAG-AAG aGAG-AAG ACA-AAA tGAA-AAA AGA-AAA ATG-AAG	Thr37Lys Ile68Lys Arg71Lys Glu116Lys Asn132Lys Glu575Lys Thr826Lys Asn1236Lys Glu1250Lys Glu1358Lys Thr1691Lys Glu1735Lys Arg1753Lys Met1775Lys	tCAG-AAG AGA-AAA AACa-AAG	Gln654Lys Arg1437Lys Asn2435Lys
BRCA1	PALB2	Breast Cancer	1	ACA-AAA ATA-AAA AGG-AAG tGAA-AAA AACc-AAA cGAA-AAA ACA-AAA AATa-AAG cGAG-AAG aGAG-AAG ACA-AAA tGAA-AAA AGA-AAA ATG-AAG	Thr37Lys Ile68Lys Arg71Lys Glu116Lys Asn132Lys Glu575Lys Thr826Lys Asn1236Lys Glu1250Lys Glu1358Lys Thr1691Lys Glu1735Lys Arg1753Lys Met1775Lys	tGAA-AAA	Glu272Lys
BRCA1	PALB2	Ovarian Cancer	1	ATG-AAG ACA-AAA ATA-AAA AGG-AAG tGAA-AAA AACc-AAA cGAA-AAA AATa-AAG	Met18Lys Thr37Lys Ile68Lys Arg71Lys Glu116Lys Asn132Lys Glu575Lys Asn1236Lys	tGAA-AAA	Glu272Lys
BRCA2	ATM	Breast Cancer	1	ACA-AAA ATA-AAA AGG-AAG tGAA-AAA AACc-AAA cGAA-AAA ACA-AAA AATa-AAG cGAG-AAG aGAG-AAG ACA-AAA tGAA-AAA AGA-AAA ATG-AAG	Thr37Lys Ile68Lys Arg71Lys Glu116Lys Asn132Lys Glu575Lys Thr826Lys Asn1236Lys Glu1250Lys Glu1358Lys Thr1691Lys Glu1735Lys Arg1753Lys Met1775Lys	tCAG-AAG AGA-AAA AACa-AAG	Gln654Lys Arg1437Lys Asn2435Lys
BRCA2	BRCA1	Breast Cancer	1	tGAA-AAA AACg-AAG AATa-AAG AGA-AAA tCAA-AAA AGA-AAA AGA-AAA aGAA-AAA aGAA-AAA	Glu534Lys Asn1878Lys Asn1880Lys Arg2268Lys Gln2384Lys Arg2488Lys Arg2659Lys Glu2981Lys Glu3002Lys	ACA-AAA ATA-AAA AGG-AAG tGAA-AAA AACc-AAA cGAA-AAA ACA-AAA AATa-AAG cGAG-AAG aGAG-AAG ACA-AAA tGAA-AAA AGA-AAA ATG-AAG	Thr37Lys Ile68Lys Arg71Lys Glu116Lys Asn132Lys Glu575Lys Thr826Lys Asn1236Lys Glu1250Lys Glu1358Lys Thr1691Lys Glu1735Lys Arg1753Lys Met1775Lys
BRCA2	BRCA1	Ovarian Cancer	1	AACg-AAG AGA-AAA AGA-AAA aGAA-AAA	Asn1878Lys Arg2488Lys Arg2659Lys Glu2981Lys	ATG-AAG ACA-AAA ATA-AAA AGG-AAG	Met18Lys Thr37Lys Ile68Lys Arg71Lys

						tGAA-AAA AACc-AAA cGAA-AAA AATa-AAG	Glu116Lys Asn132Lys Glu575Lys Asn1236Lys
BRCA2	PALB2	Breast Cancer	1	tGAA-AAA AACg-AAG AATa-AAG AGA-AAA tCAA-AAA AGA-AAA AGA-AAA aGAA-AAA aGAA-AAA	Glu534Lys Asn1878Lys Asn1880Lys Arg2268Lys Gln2384Lys Arg2488Lys Arg2659Lys Glu2981Lys Glu3002Lys	tGAA-AAA	Glu272Lys
BRCA2	PALB2	Ovarian Cancer	1	AACg-AAG AGA-AAA AGA-AAA aGAA-AAA	Asn1878Lys Arg2488Lys Arg2659Lys Glu2981Lys	tGAA-AAA	Glu272Lys
BRCA2	TP53	Breast Cancer	0	tGAA-AAA AACg-AAG AATa-AAG AGA-AAA tCAA-AAA AGA-AAA AGA-AAA aGAA-AAA aGAA-AAA	Glu534Lys Asn1878Lys Asn1880Lys Arg2268Lys Gln2384Lys Arg2488Lys Arg2659Lys Glu2981Lys Glu3002Lys	aGAG-AAG	Glu285Lys
EXO1	MLH1	Bowel Cancer	1	tGAG-AAG tGAG-AAG	Glu109Lys Glu589Lys	ATG-AAG gGAA-AAA ATG-AAG tGAG-AAG AACt-AAG cCAA-AAA tGAG-AAG AGG-AAG ATG-AAG	Met1Lys Glu23Lys Met35Lys Glu37Lys Asn38Lys Gln62Lys Glu102Lys Arg182Lys Met458Lys
EXO1	MSH2	Bowel Cancer	1	tGAG-AAG tGAG-AAG	Glu109Lys Glu589Lys	tGAA-AAA tCAG-AAG aCAG-AAG tGAA-AAA tGAA-AAA tGAA-AAA	Glu177Lys Gln409Lys Gln419Lys Glu561Lys Glu647Lys Glu749Lys
MLH3	MLH1	Bowel Cancer	1	tGAG-AAG	Glu1451Lys	ATG-AAG gGAA-AAA ATG-AAG tGAG-AAG AACt-AAG cCAA-AAA tGAG-AAG AGG-AAG ATG-AAG	Met1Lys Glu23Lys Met35Lys Glu37Lys Asn38Lys Gln62Lys Glu102Lys Arg182Lys Met458Lys
MSH2	MLH1	Bowel Cancer	1	tGAA-AAA tCAG-AAG aCAG-AAG tGAA-AAA tGAA-AAA tGAA-AAA	Glu177Lys Gln409Lys Gln419Lys Glu561Lys Glu647Lys Glu749Lys	ATG-AAG gGAA-AAA ATG-AAG tGAG-AAG AACt-AAG cCAA-AAA tGAG-AAG AGG-AAG ATG-AAG	Met1Lys Glu23Lys Met35Lys Glu37Lys Asn38Lys Gln62Lys Glu102Lys Arg182Lys Met458Lys
PTEN	TP53	Breast Cancer	0	AGA-AAA	Arg161Lys	aGAG-AAG	Glu285Lys

Methionine

Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
BRCA1	BARD1	Breast Cancer	1	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met	cGTG-ATG	Val507Met
BRCA1	UBE2I	Breast Cancer	1	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met	cGTG-ATG	Val25Met
BRCA1	ATM	Breast Cancer	0	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met	ACG-ATG	Thr1156Met
BRCA1	PALB2	Breast Cancer	1	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met	cGTG-ATG cGTG-ATG	Val932Met Val1103Met
BRCA2	BARD1	Breast Cancer	0	ATAa-ATG ACG-ATG ACG-ATG ACG-ATG ACG-ATG	Ile729Met Thr1354Met Thr1887Met Thr1915Met Thr3401Met	cGTG-ATG	Val507Met
BRCA2	ATM	Breast Cancer	0	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met	ACG-ATG	Thr1156Met
BRCA2	BRCA1	Breast Cancer	0	ATAa-ATG ACG-ATG ACG-ATG ACG-ATG ACG-ATG	Ile729Met Thr1354Met Thr1887Met Thr1915Met Thr3401Met	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met
BRCA2	BRCA1	Ovarian Cancer	0	ACG-ATG	Thr1887Met	AGG-ATG tGTG-ATG	Arg1495Met Val1534Met
BRCA2	PALB2	Breast Cancer	0	ATAa-ATG ACG-ATG ACG-ATG ACG-ATG ACG-ATG	Ile729Met Thr1354Met Thr1887Met Thr1915Met Thr3401Met	cGTG-ATG cGTG-ATG	Val932Met Val1103Met
CHEK2	BRCA2	Breast Cancer	1	ATAg-ATG ACG-ATG	Ile160Met Thr476Met	ATAa-ATG ACG-ATG ACG-ATG ACG-ATG	Ile729Met Thr1354Met Thr1915Met Thr3401Met
CHEK2	ATM	Breast Cancer	1	ATAg-ATG ACG-ATG	Ile160Met Thr476Met	ACG-ATG	Thr1156Met
CHEK2	BRCA1	Breast Cancer	0	ATAg-ATG ACG-ATG	Ile160Met Thr476Met	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met
CHEK2	BRCA1	Breast Cancer	0	ATAg-ATG ACG-ATG	Ile160Met Thr476Met	tGTG-ATG ATTc-ATG tGTG-ATG	Val271Met Ile379Met Val1534Met
MSH2	MLH1	Bowel Cancer	1	ACG-ATG ACG-ATG	Thr8Met Thr44Met	ACG-ATG cGTG-ATG	Thr117Met Val213Met

				ATTg-ATG AAG-ATG ATAa-ATG	Ile145Met Lys393Met Ile930Met	cCTG-ATG tGTG-ATG cTTG-ATG	Leu317Met Val716Met Leu724Met
MSH6	MLH1	Bowel Cancer	1	ATCt-ATG AAG-ATG ACG-ATG ACG-ATG	Ile725Met Lys854Met Thr1100Met Thr1284Met	ACG-ATG cGTG-ATG cCTG-ATG tGTG-ATG cTTG-ATG	Thr117Met Val213Met Leu317Met Val716Met Leu724Met
MSH6	MSH2	Bowel Cancer	1	ATCt-ATG AAG-ATG ACG-ATG ACG-ATG	Ile725Met Lys854Met Thr1100Met Thr1284Met	ACG-ATG ACG-ATG ATTg-ATG AAG-ATG ATAa-ATG	Thr8Met Thr44Met Ile145Met Lys393Met Ile930Met
MUTYH	MSH6	Bowel Cancer	0	tGTG-ATG cCTG-ATG	Val234Met Leu420Met	ATCt-ATG AAG-ATG ACG-ATG ACG-ATG	Ile725Met Lys854Met Thr1100Met Thr1284Met
TP53	MSH2	Bowel Cancer	0	aGTG-ATG	Val197Met	ACG-ATG ACG-ATG ATTg-ATG AAG-ATG ATAa-ATG	Thr8Met Thr44Met Ile145Met Lys393Met Ile930Met
Proline							
Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
ATM	TP53	Breast Cancer	0	aTCT-CCT	Ser707Pro	CTG-CCG CCAg-CCG	Leu188Pro Pro301Pro
BRCA1	TP53	Breast Cancer	1	CTG-CCG cTCT-CCT CAA-CCA CTT-CCT gGCA-CCA CTC-CCC CTG-CCG aTCT-CCT CGA-CCA aGCA-CCA CTA-CCA CTG-CCG CAC-CCC	Leu28Pro Ser114Pro Gln544Pro Leu623Pro Ala1277Pro Leu1407Pro Leu1564Pro Ser1577Pro Arg1751Pro Ala1752Pro Leu1764Pro Leu1780Pro His1805Pro	CTG-CCG CCAg-CCG	Leu188Pro Pro301Pro
BRCA1	ATM	Breast Cancer	1	CTG-CCG cTCT-CCT CAA-CCA CTT-CCT gGCA-CCA CTC-CCC CTG-CCG aTCT-CCT CGA-CCA aGCA-CCA CTA-CCA CTG-CCG CAC-CCC	Leu28Pro Ser114Pro Gln544Pro Leu623Pro Ala1277Pro Leu1407Pro Leu1564Pro Ser1577Pro Arg1751Pro Ala1752Pro Leu1764Pro Leu1780Pro His1805Pro	aTCT-CCT	Ser707Pro
BRCA1	PALB2	Breast Cancer	1	CTG-CCG cTCT-CCT CAA-CCA CTT-CCT gGCA-CCA CTC-CCC CTG-CCG aTCT-CCT CGA-CCA aGCA-CCA CTA-CCA CTG-CCG CAC-CCC	Leu28Pro Ser114Pro Gln544Pro Leu623Pro Ala1277Pro Leu1407Pro Leu1564Pro Ser1577Pro Arg1751Pro Ala1752Pro Leu1764Pro Leu1780Pro His1805Pro	CTT-CCT	Leu1143Pro
BRCA2	ATM	Breast Cancer	1	CTG-CCG cTCT-CCT CAA-CCA CTT-CCT gGCA-CCA CTC-CCC CTG-CCG aTCT-CCT CGA-CCA aGCA-CCA CTA-CCA CTG-CCG CAC-CCC	Leu28Pro Ser114Pro Gln544Pro Leu623Pro Ala1277Pro Leu1407Pro Leu1564Pro Ser1577Pro Arg1751Pro Ala1752Pro Leu1764Pro Leu1780Pro His1805Pro	aTCT-CCT	Ser707Pro
BRCA2	BRCA1	Breast Cancer	1	gGCT-CCT tTCA-CCA cACT-CCT CGC-CCC CTT-CCT aTCA-CCA	Ala75Pro Ser76Pro Thr582Pro Arg2336Pro Leu2653Pro Ser2835Pro	CTG-CCG cTCT-CCT CAA-CCA CTT-CCT gGCA-CCA CTC-CCC CTG-CCG aTCT-CCT CGA-CCA aGCA-CCA CTA-CCA CTG-CCG CAC-CCC	Leu28Pro Ser114Pro Gln544Pro Leu623Pro Ala1277Pro Leu1407Pro Leu1564Pro Ser1577Pro Arg1751Pro Ala1752Pro Leu1764Pro Leu1780Pro His1805Pro
BRCA2	BRCA1	Ovarian Cancer	1	tTCA-CCA CGC-CCC CTT-CCT	Ser76Pro Arg2336Pro Leu2653Pro	CTG-CCG CAA-CCA CTT-CCT CTC-CCC CGA-CCA CTA-CCA CTG-CCG CAC-CCC	Leu28Pro Gln544Pro Leu623Pro Leu1407Pro Arg1751Pro Leu1764Pro Leu1780Pro His1805Pro
BRCA2	PALB2	Breast Cancer	0	gGCT-CCT tTCA-CCA cACT-CCT CGC-CCC aTCA-CCA	Ala75Pro Ser76Pro Thr582Pro Arg2336Pro Ser2835Pro	CTT-CCT	Leu1143Pro
BRCA2	TP53	Breast Cancer	0	gGCT-CCT tTCA-CCA cACT-CCT CGC-CCC CTT-CCT aTCA-CCA	Ala75Pro Ser76Pro Thr582Pro Arg2336Pro Leu2653Pro Ser2835Pro	CTG-CCG CCAg-CCG	Leu188Pro Pro301Pro

CHEK2	BRCA2	Breast Cancer	0	CGG-CCG tACC-CCC	Arg145Pro Thr323Pro	aTCA-CCA	Ser2806Pro
MSH2	MLH1	Bowel Cancer	2	cACA-CCA CTT-CCT aGCT-CCT CTA-CCA CTG-CCG CTT-CCT CTT-CCT CTT-CCT CTT-CCT CTG-CCG CTT-CCT tACT-CCT CGT-CCT tACC-CCC aGCA-CCA CGA-CCA CTC-CCC CGA-CCA	Thr33Pro Leu93Pro Ala107Pro Leu173Pro Leu175Pro Leu187Pro Leu310Pro Leu341Pro Leu387Pro Leu421Pro Leu440Pro Thr441Pro Arg524Pro Thr552Pro Ala636Pro Arg680Pro Leu687Pro Arg711Pro	CGG-CCG CAA-CCA CGA-CCA CAT-CCT gGCT-CCT tACA-CCA aGCA-CCA cTCA-CCA aTCC-CCC CTG-CCG CAC-CCC CGT-CCT CTT-CCT CAG-CCG CTT-CCT CTC-CCC CAG-CCG CTC-CCC CTC-CCC tGCC-CCC CTT-CCT CTT-CCT tGCA-CCA CTG-CCG CGA-CCA cACT-CCT CTC-CCC aTCT-CCT CAC-CCC CTA-CCA	Arg27Pro Gln48Pro Arg100Pro His109Pro Ala111Pro Thr116Pro Ala128Pro Ser193Pro Ser247Pro Leu292Pro His329Pro Arg385Pro Leu393Pro Gln542Pro Leu549Pro Leu550Pro Gln562Pro Leu574Pro Leu582Pro Ala586Pro Leu588Pro Leu622Pro Ala623Pro Leu636Pro Arg659Pro Thr662Pro Leu676Pro Ser692Pro His718Pro Leu749Pro
MSH6	MLH1	Bowel Cancer	1	CTG-CCG gTCT-CCT CTA-CCA tGCT-CCT	Leu449Pro Ser666Pro Leu792Pro Ala1236Pro	CGG-CCG CAA-CCA aGCC-CCC CGA-CCA CAT-CCT gGCT-CCT tACA-CCA aGCA-CCA cTCA-CCA aTCC-CCC CTG-CCG CAC-CCC CGT-CCT CTT-CCT CAG-CCG CTT-CCT CTC-CCC CAG-CCG CTC-CCC CTC-CCC tGCC-CCC CTT-CCT CTT-CCT tGCA-CCA CTG-CCG CGA-CCA cACT-CCT CTC-CCC aTCT-CCT CAC-CCC CTA-CCA	Arg27Pro Gln48Pro Ala92Pro Arg100Pro His109Pro Ala111Pro Thr116Pro Ala128Pro Ser193Pro Ser247Pro Leu292Pro His329Pro Arg385Pro Leu393Pro Gln542Pro Leu549Pro Leu550Pro Gln562Pro Leu574Pro Leu582Pro Ala586Pro Leu588Pro Leu622Pro Ala623Pro Leu636Pro Arg659Pro Thr662Pro Leu676Pro Ser692Pro His718Pro Leu749Pro
MSH6	MSH2	Bowel Cancer	1	gTCT-CCT CTA-CCA tGCT-CCT	Ser666Pro Leu792Pro Ala1236Pro	cACA-CCA CTT-CCT aGCT-CCT CTA-CCA CTG-CCG CTT-CCT CTT-CCT CTT-CCT CTT-CCT CTG-CCG CTT-CCT tACT-CCT CGT-CCT tACC-CCC aGCA-CCA CGA-CCA CTC-CCC CGA-CCA	Thr33Pro Leu93Pro Ala107Pro Leu173Pro Leu175Pro Leu187Pro Leu310Pro Leu341Pro Leu387Pro Leu421Pro Leu440Pro Thr441Pro Arg524Pro Thr552Pro Ala636Pro Arg680Pro Leu687Pro Arg711Pro
MUTYH	MSH6	Bowel Cancer	0	CAA-CCA	Gln454Pro	CTG-CCG gTCT-CCT CTA-CCA tGCT-CCT	Leu449Pro Ser666Pro Leu792Pro Ala1236Pro
PMS2	MSH2	Bowel Cancer	0	CAG-CCG	Gln205Pro	cACA-CCA CTT-CCT aGCT-CCT CTA-CCA CTG-CCG CTT-CCT CTT-CCT CTT-CCT CTT-CCT CTG-CCG CTT-CCT tACT-CCT CGT-CCT tACC-CCC aGCA-CCA CGA-CCA CTC-CCC CGA-CCA	Thr33Pro Leu93Pro Ala107Pro Leu173Pro Leu175Pro Leu187Pro Leu310Pro Leu341Pro Leu387Pro Leu421Pro Leu440Pro Thr441Pro Arg524Pro Thr552Pro Ala636Pro Arg680Pro Leu687Pro Arg711Pro
PMS2	MLH1	Bowel Cancer	1	CAG-CCG	Gln205Pro	CGG-CCG CAA-CCA CGA-CCA CAT-CCT gGCT-CCT tACA-CCA aGCA-CCA cTCA-CCA aTCC-CCC CTG-CCG CAC-CCC	Arg27Pro Gln48Pro Arg100Pro His109Pro Ala111Pro Thr116Pro Ala128Pro Ser193Pro Ser247Pro Leu292Pro His329Pro

						CGT-CCT CTT-CCT CAG-CCG CTT-CCT CTC-CCC CAG-CCG CTC-CCC CTC-CCC tGCC-CCC CTT-CCT CTT-CCT tGCA-CCA CTG-CCG CGA-CCA cACT-CCT CTC-CCC aTCT-CCT CAC-CCC CTA-CCA	Arg385Pro Leu393Pro Gln542Pro Leu549Pro Leu550Pro Gln562Pro Leu574Pro Leu582Pro Ala586Pro Leu588Pro Leu622Pro Ala623Pro Leu636Pro Arg659Pro Thr662Pro Leu676Pro Ser692Pro His718Pro Leu749Pro
Termination							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	1	aAGA-TGA TGGt-TGA aGAA-TAA aGAA-TAA	Arg1039Term Trp2845Term Glu2990Term Glu642Term	cCAA-TAA tGAA-TAA TGGc-TGA TGG-TAG tGAG-TAG TGTc-TGA	Gln38Term Glu62Term Trp91Term Trp146Term Glu346Term Cys277Term
CHEK2	BRCA2	Prostate Cancer	0	cGAG-TAG	Glu239Term	gCAA-TAA TACa-TAA cAAA-TAA TATc-TAA	Gln84Term Tyr1135Term Lys1489Term Tyr1762Term
CHEK2	BRCA2	Breast Cancer	2	aGAG-TAG tCGA-TGA aGAA-TAA cCAG-TAG	Glu64Term Arg95Term Glu275Term Gln510Term	tGGA-TGA TTA-TAA TGG-TAG TGGt-TGA tGAA-TAA aGAA-TAA tGAA-TAA aGAA-TAA TACg-TAG cGAA-TAA aCAA-TAA cAAA-TAA TACc-TAG aGAA-TAA TTA-TGA TGTc-TGA aCAA-TAA TGG-TAG tCAA-TAA aAGA-TGA TCA-TAA aAAA-TAA aCAA-TAA cAAG-TAG tGAA-TAA TTG-TAG cCAA-TAA TCA-TAA tAAG-TAG tGAA-TAA aGAA-TAA aGAA-TAA TTA-TGA TGGc-TGA TTG-TAG TCA-TAA cCAG-TAG aCAA-TAA TCA-TGA TCA-TAA TACa-TAA tCAA-TAA TCA-TGA TCA-TAA TGTg-TGA tAAA-TAA TTA-TGA aGAA-TAA TGGg-TGA TTA-TAA gAAG-TAG TTA-TGA tCAA-TAA tCAG-TAG gCAG-TAG TCA-TGA TCA-TGA cCAG-TAG TTA-TGA TCA-TGA cCAA-TAA aGAA-TAA tGAA-TAA TTA-TAA aCAG-TAG TATg-TAG aGAA-TAA TCA-TGA TACa-TAA tCAA-TAA aGAA-TAA TCA-TGA TCA-TGA gGAA-TAA tGAA-TAA gGAA-TAA TCA-TAA TCA-TGA cCAA-TAA tGAA-TAA TGTa-TGA	Gly4Term Leu24Term Trp31Term Trp31Term Glu33Term Glu34Term Glu45Term Glu49Term Tyr57Term Glu58Term Gln66Term Lys82Term Tyr91Term Glu97Term Leu105Term Cys132Term Gln147Term Trp194Term Gln258Term Arg259Term Ser273Term Lys318Term Gln321Term Lys385Term Glu394Term Leu414Term Gln421Term Ser442Term Lys467Term Glu475Term Glu510Term Glu532Term Leu557Term Trp563Term Leu583Term Ser611Term Gln619Term Gln742Term Ser744Term Ser791Term Tyr792Term Gln819Term Ser846Term Ser871Term Cys916Term Lys944Term Leu971Term Glu984Term Trp993Term Leu997Term Lys1026Term Leu1053Term Gln1056Term Gln1089Term Gln1095Term Ser1099Term Ser1121Term Gln1148Term Leu1152Term Ser1262Term Gln1291Term Glu1308Term Glu1320Term Leu1334Term Gln1408Term Tyr1480Term Glu1518Term Ser1630Term Tyr1655Term Gln1701Term Glu1703Term Ser1760Term Ser1764Term Glu1812Term Glu1857Term Glu1876Term Ser1882Term Ser1882Term Gln1886Term Glu1912Term Cys1913Term

						aGAA-TAA TTA-TAA gGAA-TAA TCA-TAA TCA-TGA TCA-TAA cCAG-TAG aCAA-TAA aCAA-TAA cAAA-TAA cGAA-TAA TCA-TGA aGAA-TAA cCAA-TAA tGGA-TGA TTA-TGA TCA-TGA tCAA-TAA aGAA-TAA TCA-TGA tGAA-TAA tAAA-TAA TCA-TAA tCGA-TGA tCAA-TAA tCAA-TAA TCA-TGA gAGA-TGA tGAA-TAA aCAG-TAG gCAA-TAA cAAA-TAA tCAG-TAG aCAG-TAG gCGA-TGA tCGA-TGA tAAA-TAA aCAG-TAG TGG-TAG TGGc-TGA TGG-TAG TATa-TAA TGGa-TGA TGTg-TGA tCAA-TAA TATg-TAG cAGA-TGA TCG-TAG TGTg-TGA TCA-TGA cCAA-TAA TGGt-TGA cCAG-TAG TTA-TGA aGAA-TAA TTA-TGA TGG-TAG cCAA-TAA aCAA-TAA tGAA-TAA aAAA-TAA aCAG-TAG gCAA-TAA gCAA-TAA aCAA-TAA aCAA-TAA aCAA-TAA TCA-TAA TCA-TGA TCA-TAA TATt-TAG aGGA-TGA TACa-TAA aCAG-TAG aCAA-TAA gAAA-TAA TATt-TAG cGAA-TAA TACa-TAA TACa-TAG TGG-TAG tGAG-TAG cCAG-TAG gCGA-TGA tCAA-TAA TGG-TAG TACg-TAG cGAA-TAA tAAA-TAA aCGA-TGA	Glu1928Term Leu1930Term Glu1953Term Ser1955Term Ser1955Term Ser1970Term Gln1987Term Gln1994Term Gln1998Term Lys2013Term Glu2020Term Ser2022Term Glu2028Term Gln2042Term Gly2057Term Leu2080Term Ser2120Term Gln2157Term Glu2183Term Ser2219Term Glu2239Term Lys2244Term Ser2267Term Arg2318Term Gln2342Term Gln2354Term Ser2378Term Arg2394Term Glu2420Term Gln2421Term Gln2435Term Lys2459Term Gln2485Term Gln2491Term Arg2494Term Arg2520Term Lys2538Term Gln2580Term Trp2586Term Trp2586Term Trp2619Term Tyr2621Term Trp2626Term Cys2636Term Gln2655Term Tyr2660Term Arg2668Term Ser2670Term Cys2689Term Ser2695Term Gln2714Term Trp2725Term Gln2731Term Leu2732Term Glu2772Term Leu2776Term Trp2788Term Gln2858Term Gln2859Term Glu2874Term Lys2882Term Gln2893Term Gln2894Term Gln2899Term Gln2941Term Gln2957Term Gln2960Term Ser2978Term Ser2984Term Ser2993Term Tyr2997Term Gly3003Term Tyr3006Term Gln3026Term Gln3037Term Lys3083Term Tyr3092Term Glu3096Term Tyr3098Term Tyr3098Term Trp3106Term Glu3111Term Gln3126Term Arg3128Term Gln3156Term Trp3191Term Tyr3308Term Glu3309Term Lys3326Term Arg3384Term
CHEK2	ATM	Breast Cancer	1	aGAG-TAG tCGA-TGA aGAA-TAA cCAG-TAG	Glu64Term Arg95Term Glu275Term Gln510Term	aAGA-TGA TGGt-TGA aGAA-TAA aGAA-TAA	Arg1039Term Trp2845Term Glu2990Term Glu642Term
CHEK2	TP53	Breast Cancer	1	aGAG-TAG tCGA-TGA aGAA-TAA cCAG-TAG	Glu64Term Arg95Term Glu275Term Gln510Term	cCAA-TAA tGAA-TAA TGGc-TGA TGG-TAG tGAG-TAG TGTc-TGA	Gln38Term Glu62Term Trp91Term Trp146Term Glu346Term Cys277Term
CHEK2	BRCA1	Breast Cancer	2	aGAG-TAG tCGA-TGA aGAA-TAA cCAG-TAG	Glu64Term Arg95Term Glu275Term Gln510Term	TTA-TGA aCAA-TAA gCAG-TAG gGAG-TAG TTG-TAG aCAG-TAG TTA-TAA aCAA-TAA tCAA-TAA TATg-TAG TACa-TAA cGAA-TAA cCAA-TAA gCAG-TAG	Leu3Term Gln12Term Gln19Term Glu29Term Leu30Term Gln60Term Leu63Term Gln74Term Gln81Term Tyr101Term Tyr130Term Glu143Term Gln155Term Gln169Term

						aCAA-TAA	Gln206Term
						TGTg-TGA	Cys226Term
						tGAG-TAG	Glu255Term
						TATc-TAG	Tyr261Term
						TCA-TGA	Ser267Term
						TTA-TGA	Leu283Term
						aCAG-TAG	Gln284Term
						TGG-TAG	Trp321Term
						aAAA-TAA	Lys338Term
						TGG-TAG	Trp353Term
						tAAG-TAG	Lys355Term
						gCAG-TAG	Gln356Term
						TCA-TGA	Ser361Term
						TGG-TAG	Trp372Term
						TGGa-TGA	Trp372Term
						tCAG-TAG	Gln380Term
						gAAA-TAA	Lys381Term
						TCA-TAA	Ser398Term
						gGAG-TAG	Glu402Term
						tGAG-TAG	Glu418Term
						TATt-TAG	Tyr422Term
						aGAG-TAG	Glu427Term
						TTA-TGA	Leu431Term
						TGTa-TGA	Cys442Term
						tGAA-TAA	Glu445Term
						TCA-TAA	Ser451Term
						gAAG-TAG	Lys467Term
						TTA-TGA	Leu474Term
						aGGA-TGA	Gly484Term
						tGAG-TAG	Glu489Term
						aCAG-TAG	Gln491Term
						aCAA-TAA	Gln494Term
						tAAA-TAA	Lys505Term
						gAGA-TGA	Arg507Term
						TCA-TGA	Ser510Term
						tCAA-TAA	Gln526Term
						tCAG-TAG	Gln534Term
						cCAA-TAA	Gln538Term
						gCAG-TAG	Gln541Term
						tCAA-TAA	Gln544Term
						tGAG-TAG	Glu554Term
						tCAG-TAG	Gln563Term
						aGAA-TAA	Glu577Term
						cAAA-TAA	Lys581Term
						TTA-TGA	Leu598Term
						TCA-TGA	Ser603Term
						gAAG-TAG	Lys614Term
						tGAA-TAA	Glu638Term
						TTG-TAG	Leu639Term
						aGAG-TAG	Glu649Term
						aAAG-TAG	Lys654Term
						cCAA-TAA	Gln657Term
						tAAA-TAA	Lys672Term
						cAAG-TAG	Lys679Term
						aCAG-TAG	Gln687Term
						TCA-TGA	Ser713Term
						aGAA-TAA	Glu732Term
						aGAG-TAG	Glu733Term
						aGAA-TAA	Glu755Term
						TCA-TAA	Ser770Term
						tCAG-TAG	Gln780Term
						TTA-TAA	Leu785Term
						aGAA-TAA	Glu797Term
						cAAG-TAG	Lys812Term
						gGGA-TGA	Gly813Term
						tCAG-TAG	Gln855Term
						TATt-TAG	Tyr856Term
						gCAG-TAG	Gln858Term
						TCA-TGA	Ser864Term
						TCA-TAA	Ser868Term
						TCA-TGA	Ser868Term
						aGAA-TAA	Glu879Term
						gGAA-TAA	Glu881Term
						aCAA-TAA	Gln895Term
						tGAA-TAA	Glu904Term
						aCAA-TAA	Gln905Term
						aGAA-TAA	Glu908Term
						tCAA-TAA	Gln910Term
						aCAG-TAG	Gln921Term
						TCA-TAA	Ser955Term
						tGGA-TGA	Gly972Term
						TTA-TAA	Leu974Term
						aCAA-TAA	Gln975Term
						TATc-TAG	Tyr978Term
						TCA-TAA	Ser988Term
						TCA-TGA	Ser1007Term
						aGAA-TAA	Glu1038Term
						TCA-TGA	Ser1041Term
						tGAA-TAA	Glu1060Term
						TTG-TAG	Leu1080Term
						TTA-TGA	Leu1086Term
						gCAA-TAA	Gln1090Term
						tGAA-TAA	Glu1107Term
						TATg-TAG	Tyr1113Term
						tGAA-TAA	Glu1114Term
						TCA-TGA	Ser1130Term
						TTA-TGA	Leu1133Term
						aGAA-TAA	Glu1134Term
						aCAG-TAG	Gln1135Term
						tCAG-TAG	Gln1144Term
						tGAA-TAA	Glu1158Term
						gGAA-TAA	Glu1172Term
						aGAG-TAG	Glu1185Term
						TTG-TAG	Leu1198Term
						tCAG-TAG	Gln1200Term
						cCGA-TGA	Arg1203Term
						aGAG-TAG	Glu1214Term
						tGAA-TAA	Glu1221Term
						cGAG-TAG	Glu1250Term
						TGTc-TGA	Cys1251Term
						gGAG-TAG	Glu1258Term
						TTA-TGA	Leu1261Term
						TCA-TAA	Ser1262Term
						cCAG-TAG	Gln1273Term
						tCAG-TAG	Gln1281Term

						TCA-TAA aCAG-TAG cCAG-TAG aCAA-TAA cCAG-TAG TTG-TAG aGAA-TAA gCAA-TAA TGTg-TGA tCAG-TAG gCAG-TAG gCAA-TAA cCAG-TAG gGAA-TAA tGAA-TAA TTA-TGA aCAG-TAG TACc-TAA gCGA-TGA aCAA-TAA aCAG-TAG cCAG-TAG gGAA-TAA TCA-TGA TCA-TAA TGGt-TGA gGAG-TAG gCAA-TAA aCAG-TAG aGAG-TAG gGGA-TGA TACc-TAG tGAA-TAA TCA-TGA gAAA-TAA cCAA-TAA aGAA-TAA aGAA-TAA cAAG-TAG gAAA-TAA tGAG-TAG TATt-TAG aGGA-TGA TGGg-TGA TATt-TAG TGG-TAG TGGg-TGA aGAA-TAA aAAA-TAA gCGA-TGA aGAA-TAA cCAG-TAG aGAA-TAA TATg-TAA tCAA-TAA gGAA-TAA TGG-TAG TGGa-TGA aCAG-TAG gAAG-TAG gCAG-TAG TGG-TAG TGGa-TGA aGAG-TAG cCGA-TGA aGAG-TAG TGG-TAG TGGg-TGA cCAG-TAG TGCc-TGA TACc-TAG	Ser1298Term Gln1299Term Gln1313Term Gln1323Term Gln1331Term Leu1351Term Glu1353Term Gln1359Term Cys1372Term Gln1395Term Gln1396Term Gln1401Term Gln1408Term Glu1410Term Glu1413Term Leu1418Term Gln1420Term Tyr1429Term Arg1443Term Gln1447Term Gln1458Term Gln1467Term Glu1494Term Ser1496Term Ser1503Term Trp1508Term Glu1535Term Gln1537Term Gln1538Term Glu1541Term Gly1560Term Tyr1563Term Glu1581Term Ser1587Term Lys1601Term Gln1604Term Glu1644Term Glu1661Term Lys1667Term Lys1690Term Glu1694Term Tyr1703Term Gly1710Term Trp1712Term Tyr1716Term Trp1718Term Trp1718Term Glu1725Term Lys1727Term Arg1751Term Glu1754Term Gln1756Term Glu1765Term Tyr1769Term Gln1779Term Glu1781Term Trp1782Term Trp1782Term Gln1785Term Lys1793Term Gln1811Term Trp1815Term Trp1815Term Glu1817Term Arg1835Term Glu1836Term Trp1837Term Trp1837Term Gln1846Term Cys1847Term Tyr1853Term
MUTYH	MSH6	Bowel Cancer	1	TATg-TAA gGAG-TAG gCAG-TAG aCAG-TAG cCGA-TGA TACg-TAA cCAG-TAG gGAA-TAA	Tyr128Term Glu196Term Gln314Term Gln338Term Arg97Term Tyr104Term Gln267Term Glu480Term	aCAG-TAG cGGA-TGA TGGg-TGA TCA-TGA TACg-TAA TACg-TAG aCAG-TAG aCAA-TAA gCGA-TGA TCA-TAA gGAA-TAA tCAA-TAA TATg-TAA aCGA-TGA aCAG-TAG gGAG-TAG aCGA-TGA TCA-TAA TCA-TGA TGTg-TGA TCA-TAA aCGA-TGA tCAG-TAG TTG-TAG tCGA-TGA cCAA-TAA TACc-TAG cGAG-TAG aCGA-TGA aGAA-TAA gCGA-TGA gCAG-TAG tCGA-TGA gGAA-TAA TACg-TAA TATt-TAG tCAA-TAA cCAG-TAG aCGA-TGA	Gln4Term Gly74Term Trp142Term Ser200Term Tyr214Term Tyr214Term Gln232Term Gln236Term Arg240Term Ser252Term Glu272Term Gln344Term Tyr433Term Arg482Term Gln485Term Glu493Term Arg495Term Ser612Term Ser664Term Cys687Term Ser702Term Arg732Term Gln835Term Leu905Term Arg922Term Gln939Term Tyr977Term Glu995Term Arg1005Term Glu1023Term Arg1035Term Gln1048Term Arg1068Term Glu1119Term Tyr1159Term Tyr1256Term Gln1258Term Gln1280Term Arg1331Term
PMS1	MLH1	Bowel Cancer	1	tCAG-TAG	Gln233Term	TCG-TAG cGAG-TAG gGAA-TAA cCAG-TAG	Ser2Term Glu13Term Glu23Term Gln26Term

						tGAG-TAG cCAA-TAA aGAA-TAA gCAG-TAG tGAG-TAG TATg-TAG tCGA-TGA tGGA-TGA aAAG-TAG TACa-TAA TACa-TAG TCA-TAA tGGA-TGA tCAA-TAA cCAG-TAG TTA-TAA TCA-TGA aAAA-TAA tCGA-TGA TACt-TAG TCA-TAA TCA-TGA TATg-TAA TTG-TAG TTA-TGA aGAA-TAA cCAG-TAG cGAG-TAG gCAG-TAG cCAG-TAG aCAG-TAG gCAG-TAG cAAA-TAA cCAG-TAG gCAG-TAG TCA-TGA gAAG-TAG cAGA-TGA cCGA-TGA cCAG-TAG aGAA-TAA gCAG-TAG gGGA-TGA tCAG-TAG TGGg-TGA aCAG-TAG TTA-TAA TACc-TAA tGAA-TAA aGAA-TAA TACc-TAG cCAG-TAG cAAA-TAA gAAG-TAG TATt-TAG tCGA-TGA TGGg-TGA TGCg-TGA TATt-TAG TCA-TGA cCAG-TAG gCAG-TAG TGG-TAG TGGa-TGA TGG-TAG TGGa-TGA TATa-TAA tAAA-TAA aGAA-TAA gCAG-TAG TACa-TAG	Glu37Term Gln62Term Glu71Term Gln86Term Glu89Term Tyr97Term Arg100Term Gly122Term Lys123Term Tyr126Term Tyr126Term Ser131Term Gly133Term Gln146Term Gln149Term Leu166Term Ser193Term Lys196Term Arg226Term Tyr251Term Ser252Term Ser269Term Tyr280Term Leu284Term Leu296Term Glu297Term Gln301Term Glu319Term Gln327Term Gln382Term Gln391Term Gln398Term Lys402Term Gln409Term Gln426Term Ser459Term Lys461Term Arg470Term Arg487Term Gln510Term Glu512Term Gln516Term Gly517Term Gln537Term Trp538Term Gln542Term Leu547Term Tyr548Term Glu557Term Glu558Term Tyr561Term Gln562Term Lys604Term Lys618Term Tyr625Term Arg659Term Trp666Term Cys680Term Tyr684Term Ser698Term Gln700Term Gln701Term Trp712Term Trp712Term Trp714Term Trp714Term Tyr721Term Lys732Term Glu736Term Gln742Term Tyr750Term
PMS2	MSH2	Bowel Cancer	1	TGTg-TGA gCAG-TAG gCAG-TAG tCGA-TGA cCAG-TAG tCGA-TGA tAAA-TAA gAAA-TAA aCGA-TGA gAAG-TAG cGAA-TAA	Cys73Term Gln233Term Gln235Term Arg315Term Gln317Term Arg563Term Lys580Term Lys614Term Arg628Term Lys647Term Glu661Term	gCAG-TAG gGAG-TAG cGAG-TAG gGAG-TAG cCAG-TAG TACa-TAG gCAG-TAG tGAA-TAA TATa-TAG tGAA-TAA TCA-TGA cCAG-TAG aCAG-TAG tGAA-TAA tCAG-TAG cCAG-TAG aGGA-TGA aGAG-TAG aCAG-TAG tCAG-TAG gAAA-TAA gCAG-TAG TTA-TGA aGAA-TAA TCA-TAA TCA-TGA tGGA-TGA aCAG-TAG tGAA-TAA cCAG-TAG gAAA-TAA TTG-TAG tCAG-TAG tGAA-TAA tCAG-TAG tAAA-TAA tCAA-TAA aGGA-TGA cCAG-TAG TGG-TAG TGGa-TGA aGAG-TAG gAGA-TGA	Gln10Term Glu28Term Glu48Term Glu56Term Gln61Term Tyr66Term Gln76Term Glu101Term Tyr121Term Glu132Term Ser142Term Gln158Term Gln170Term Glu177Term Gln183Term Gln193Term Gly204Term Glu205Term Gln215Term Gln239Term Lys246Term Gln252Term Leu277Term Glu278Term Ser281Term Ser281Term Gly287Term Gln288Term Glu290Term Gln298Term Lys301Term Leu302Term Gln314Term Glu318Term Gln324Term Lys334Term Gln337Term Gly338Term Gln344Term Trp345Term Trp345Term Glu357Term Arg359Term

						gGAA-TAA gCAG-TAG aCAA-TAA tCGA-TGA cCGA-TGA tCAA-TAA aCAA-TAA aCAA-TAA TACc-TAA cCGA-TGA tCAA-TAA aCAG-TAG gGAA-TAA cCAG-TAG gAAA-TAA TTA-TAA cAAG-TAG tCAG-TAG gGAA-TAA TTA-TGA tGAA-TAA TCA-TGA aAGA-TGA aGAA-TAA aAAG-TAG gCAG-TAG TTA-TGA aCAG-TAG TACt-TAG tAAA-TAA cCAG-TAG tGAA-TAA tAAA-TAA TATg-TAG cCAG-TAG TATg-TAG gCAG-TAG tCAG-TAG TCA-TGA TATg-TAG aCGA-TGA aCAA-TAA TACt-TAG aCAG-TAG TATa-TAA tCGA-TGA TGTg-TGA tGAG-TAG cCGA-TGA tCAA-TAA gGAA-TAA TCA-TAA TTA-TGA TACg-TAG tGGA-TGA TGGg-TGA TGCa-TGA tGAA-TAA TTA-TGA tCAG-TAG tCAA-TAA aCAG-TAG tGAG-TAG aGAA-TAA TCG-TAG gCAA-TAA gCAA-TAA tCAG-TAG tGAA-TAA aCGA-TGA	Glu364Term Gln374Term Gln377Term Arg383Term Arg389Term Gln395Term Gln397Term Gln402Term Tyr405Term Arg406Term Gln413Term Gln419Term Glu422Term Gln429Term Lys430Term Leu431Term Lys449Term Gln451Term Glu452Term Leu458Term Glu467Term Ser473Term Arg482Term Glu483Term Lys490Term Gln493Term Leu496Term Gln518Term Tyr522Term Lys537Term Gln545Term Glu561Term Lys567Term Tyr570Term Gln574Term Tyr588Term Gln593Term Gln601Term Ser612Term Tyr619Term Arg621Term Gln629Term Tyr656Term Gln662Term Tyr678Term Arg680Term Cys697Term Glu698Term Arg711Term Gln718Term Glu731Term Ser743Term Leu744Term Tyr757Term Gly759Term Trp764Term Cys778Term Glu808Term Leu811Term Gln816Term Gln824Term Gln846Term Glu852Term Glu859Term Ser860Term Gln861Term Gln879Term Gln885Term Glu914Term Arg929Term
PMS2	MLH1	Bowel Cancer	1	TGTg-TGA gCAG-TAG gCAG-TAG tCGA-TGA cCAG-TAG tCGA-TGA tAAA-TAA gAAA-TAA aCGA-TGA gAAG-TAG cGAA-TAA	Cys73Term Gln233Term Gln235Term Arg315Term Gln317Term Arg563Term Lys580Term Lys614Term Arg628Term Lys647Term Glu661Term	TCG-TAG cGAG-TAG gGAA-TAA cCAG-TAG tGAG-TAG cCAA-TAA aGAA-TAA gCAG-TAG tGAG-TAG TATg-TAG tCGA-TGA tGGA-TGA aAAG-TAG TACa-TAA TACa-TAG TCA-TAA tGGA-TGA tCAA-TAA cCAG-TAG TTA-TAA TCA-TGA aAAA-TAA tCGA-TGA TACt-TAG TCA-TAA TCA-TGA TATg-TAA TTG-TAG TTA-TGA aGAA-TAA cCAG-TAG cGAG-TAG gCAG-TAG cCAG-TAG aCAG-TAG gCAG-TAG cAAA-TAA cCAG-TAG gCAG-TAG TCA-TGA gAAG-TAG cAGA-TGA cCGA-TGA cCAG-TAG	Ser2Term Glu13Term Glu23Term Gln26Term Glu37Term Gln62Term Glu71Term Gln86Term Glu89Term Tyr97Term Arg100Term Gly122Term Lys123Term Tyr126Term Tyr126Term Ser131Term Gly133Term Gln146Term Gln149Term Leu166Term Ser193Term Lys196Term Arg226Term Tyr251Term Ser252Term Ser269Term Tyr280Term Leu284Term Leu296Term Glu297Term Gln301Term Glu319Term Gln327Term Gln382Term Gln391Term Gln398Term Lys402Term Gln409Term Gln426Term Ser459Term Lys461Term Arg470Term Arg487Term Gln510Term

						aGAA-TAA gCAG-TAG gGGA-TGA tCAG-TAG TGGg-TGA aCAG-TAG TTA-TAA TACc-TAA tGAA-TAA aGAA-TAA TACc-TAG cCAG-TAG cAAA-TAA gAAG-TAG TATt-TAG tCGA-TGA TGGg-TGA TGCg-TGA TATt-TAG TCA-TGA cCAG-TAG gCAG-TAG TGG-TAG TGGa-TGA TGG-TAG TGGa-TGA TATa-TAA tAAA-TAA aGAA-TAA gCAG-TAG TACa-TAG	Glu512Term Gln516Term Gly517Term Gln537Term Trp538Term Gln542Term Leu547Term Tyr548Term Glu557Term Glu558Term Tyr561Term Gln562Term Lys604Term Lys618Term Tyr625Term Arg659Term Trp666Term Cys680Term Tyr684Term Ser698Term Gln700Term Gln701Term Trp712Term Trp712Term Trp714Term Trp714Term Tyr721Term Lys732Term Glu736Term Gln742Term Tyr750Term
TP53	BARD1	Breast Cancer	0	cCAA-TAA tGAA-TAA TGGc-TGA TGG-TAG tGAG-TAG TGTc-TGA	Gln38Term Glu62Term Trp91Term Trp146Term Glu346Term Cys277Term	gCAG-TAG	Gln564Term
XRCC2	RAD51C	Breast Cancer	1	cCGA-TGA	Arg17Term	aCAA-TAA cCGA-TGA aCAA-TAA cCGA-TGA	Gln133Term Arg193Term Gln222Term Arg319Term

Threonine

Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
BRCA1	BARD1	Breast Cancer	1	ATG-ACG ATT-ACT ATG-ACG tCCC-ACC AAA-ACA AAG-ACG ATA-ACA ATT-ACT ATG-ACG ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Met1Thr Ile15Thr Met18Thr Pro25Thr Lys45Thr Lys110Thr Ile845Thr Ile1391Thr Met1628Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr	ATA-ACA	Ile509Thr
BRCA1	ATM	Breast Cancer	1	ATG-ACG ATT-ACT ATG-ACG tCCC-ACC AAA-ACA AAG-ACG ATA-ACA ATT-ACT ATG-ACG ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Met1Thr Ile15Thr Met18Thr Pro25Thr Lys45Thr Lys110Thr Ile845Thr Ile1391Thr Met1628Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr	ATG-ACG aGCT-ACT ATA-ACA AGG-ACG gGCA-ACA ATT-ACT ATG-ACG ATG-ACG ATT-ACT	Met779Thr Ala1427Thr Ile1525Thr Arg2105Thr Ala2274Thr Ile2401Thr Met2531Thr Met2667Thr Ile2914Thr
BRCA1	PALB2	Breast Cancer	1	ATG-ACG ATT-ACT ATG-ACG tCCC-ACC AAA-ACA AAG-ACG ATA-ACA ATT-ACT ATG-ACG ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Met1Thr Ile15Thr Met18Thr Pro25Thr Lys45Thr Lys110Thr Ile845Thr Ile1391Thr Met1628Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr	ATA-ACA	Ile1180Thr
BRCA2	BARD1	Breast Cancer	1	cGCT-ACT gGCT-ACT ATA-ACA aCCC-ACC AAA-ACA ATA-ACA AGG-ACG AGA-ACA aGCC-ACC tGCT-ACT gGCC-ACC ATG-ACG	Ala248Thr Ala495Thr Ile505Thr Pro920Thr Lys1445Thr Ile2490Thr Arg2602Thr Arg2659Thr Ala2770Thr Ala2786Thr Ala2951Thr Met3118Thr	ATA-ACA	Ile509Thr
BRCA2	ATM	Breast Cancer	1	ATG-ACG ATT-ACT ATG-ACG tCCC-ACC AAA-ACA AAG-ACG ATA-ACA ATT-ACT ATG-ACG ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Met1Thr Ile15Thr Met18Thr Pro25Thr Lys45Thr Lys110Thr Ile845Thr Ile1391Thr Met1628Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr	ATG-ACG aGCT-ACT ATA-ACA AGG-ACG gGCA-ACA ATT-ACT ATG-ACG ATG-ACG ATT-ACT	Met779Thr Ala1427Thr Ile1525Thr Arg2105Thr Ala2274Thr Ile2401Thr Met2531Thr Met2667Thr Ile2914Thr
BRCA2	BRCA1	Breast Cancer	1	cGCT-ACT gGCT-ACT ATA-ACA aCCC-ACC	Ala248Thr Ala495Thr Ile505Thr Pro920Thr	ATG-ACG ATT-ACT ATG-ACG tCCC-ACC	Met1Thr Ile15Thr Met18Thr Pro25Thr

				AAA-ACA ATA-ACA AGG-ACG AGA-ACA aGCC-ACC tGCT-ACT gGCC-ACC ATG-ACG	Lys1445Thr Ile2490Thr Arg2602Thr Arg2659Thr Ala2770Thr Ala2786Thr Ala2951Thr Met3118Thr	AAA-ACA AAG-ACG ATA-ACA ATT-ACT ATG-ACG ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Lys45Thr Lys110Thr Ile845Thr Ile1391Thr Met1628Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr
BRCA2	BRCA1	Ovarian Cancer	2	ATG-ACG gGCT-ACT aCCC-ACC AGG-ACG AGA-ACA aGCC-ACC	Met1Thr Ala495Thr Pro920Thr Arg2602Thr Arg2659Thr Ala2770Thr	ATG-ACG ATT-ACT tCCC-ACC AAA-ACA ATA-ACA ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Met1Thr Ile15Thr Pro25Thr Lys45Thr Ile845Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr
BRCA2	PALB2	Breast Cancer	1	cGCT-ACT gGCT-ACT ATA-ACA aCCC-ACC AAA-ACA ATA-ACA AGG-ACG AGA-ACA aGCC-ACC tGCT-ACT gGCC-ACC ATG-ACG	Ala248Thr Ala495Thr Ile505Thr Pro920Thr Lys1445Thr Ile2490Thr Arg2602Thr Arg2659Thr Ala2770Thr Ala2786Thr Ala2951Thr Met3118Thr	ATA-ACA	Ile1180Thr
CHEK2	ATM	Breast Cancer	1	ATA-ACA	Ile364Thr	ATG-ACG aGCT-ACT ATA-ACA AGG-ACG gGCA-ACA ATT-ACT ATG-ACG ATG-ACG ATT-ACT	Met779Thr Ala1427Thr Ile1525Thr Arg2105Thr Ala2274Thr Ile2401Thr Met2531Thr Met2667Thr Ile2914Thr
CHEK2	BRCA1	Breast Cancer	1	ATA-ACA	Ile364Thr	ATG-ACG ATT-ACT ATG-ACG tCCC-ACC AAA-ACA AAG-ACG ATA-ACA ATT-ACT ATG-ACG ATG-ACG ATG-ACG tGCT-ACT tGCA-ACA	Met1Thr Ile15Thr Met18Thr Pro25Thr Lys45Thr Lys110Thr Ile845Thr Ile1391Thr Met1628Thr Met1652Thr Met1783Thr Ala1789Thr Ala1823Thr
MLH3	MLH1	Bowel Cancer	1	tGCT-ACT	Ala1394Thr	ATG-ACG ATC-ACC AGT-ACT gGCT-ACT tGCA-ACA AAC-ACC AAG-ACG aCCC-ACC ATC-ACC cGCT-ACT	Met1Thr Ile25Thr Ser295Thr Ala441Thr Ala492Thr Asn551Thr Lys618Thr Pro640Thr Ile655Thr Ala681Thr
MSH2	MLH1	Bowel Cancer	1	gGCG-ACG aGCA-ACA tGCA-ACA ATT-ACT aCCA-ACA ATT-ACT tGCT-ACT AAA-ACA	Ala2Thr Ala72Thr Ala305Thr Ile577Thr Pro622Thr Ile691Thr Ala834Thr Lys931Thr	ATG-ACG ATC-ACC AGT-ACT gGCT-ACT tGCA-ACA AAC-ACC AAG-ACG aCCC-ACC ATC-ACC cGCT-ACT	Met1Thr Ile25Thr Ser295Thr Ala441Thr Ala492Thr Asn551Thr Lys618Thr Pro640Thr Ile655Thr Ala681Thr
MSH6	MLH1	Bowel Cancer	1	cCCC-ACC ATA-ACA	Pro1087Thr Ile1115Thr	ATG-ACG ATC-ACC AGT-ACT gGCT-ACT tGCA-ACA AAC-ACC AAG-ACG aCCC-ACC ATC-ACC cGCT-ACT	Met1Thr Ile25Thr Ser295Thr Ala441Thr Ala492Thr Asn551Thr Lys618Thr Pro640Thr Ile655Thr Ala681Thr
MSH6	MSH2	Bowel Cancer	0	cCCC-ACC ATA-ACA	Pro1087Thr Ile1115Thr	gGCG-ACG aGCA-ACA tGCA-ACA ATT-ACT aCCA-ACA ATT-ACT tGCT-ACT AAA-ACA	Ala2Thr Ala72Thr Ala305Thr Ile577Thr Pro622Thr Ile691Thr Ala834Thr Lys931Thr
MUTYH	MSH6	Bowel Cancer	0	tGCT-ACT ACGc-ACC	Ala473Thr Thr477Thr	cCCC-ACC ATA-ACA	Pro1087Thr Ile1115Thr
TP53	MSH2	Bowel Cancer	0	ATC-ACC	Ile195Thr	gGCG-ACG aGCA-ACA tGCA-ACA ATT-ACT aCCA-ACA ATT-ACT tGCT-ACT AAA-ACA	Ala2Thr Ala72Thr Ala305Thr Ile577Thr Pro622Thr Ile691Thr Ala834Thr Lys931Thr
PMS1	MLH1	Bowel Cancer	1	ATG-ACG	Met394Thr	ATG-ACG ATC-ACC AGT-ACT gGCT-ACT tGCA-ACA AAC-ACC AAG-ACG aCCC-ACC ATC-ACC cGCT-ACT	Met1Thr Ile25Thr Ser295Thr Ala441Thr Ala492Thr Asn551Thr Lys618Thr Pro640Thr Ile655Thr Ala681Thr
Tyrosine							

Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
BRCA1	ATM	Breast Cancer	1	TGT-TAT TGT-TAT TGC-TAC TGC-TAC TGT-TAT TGT-TAT tGAT-TAT tAAC-TAC gCAC-TAC TCT-TAT gCAT-TAT aGAT-TAT	Cys24Tyr Cys39Tyr Cys44Tyr Cys47Tyr Cys61Tyr Cys64Tyr Asp67Tyr Asn158Tyr His662Tyr Ser1253Tyr His1421Tyr Asp1739Tyr	TGT-TAT tCAT-TAT tCAT-TAT aGAT-TAT	Cys107Tyr His231Tyr His1380Tyr Asp1682Tyr
BRCA2	ATM	Breast Cancer	1	TGT-TAT TGT-TAT TGC-TAC TGC-TAC TGT-TAT TGT-TAT tGAT-TAT tAAC-TAC tCAC-TAC gCAC-TAC TCT-TAT gCAT-TAT aGAT-TAT aGAT-TAT	Cys24Tyr Cys39Tyr Cys44Tyr Cys47Tyr Cys61Tyr Cys64Tyr Asp67Tyr Asn158Tyr His448Tyr His662Tyr Ser1253Tyr His1421Tyr Asp1739Tyr Asp1778Tyr	TGT-TAT tCAT-TAT tCAT-TAT aGAT-TAT	Cys107Tyr His231Tyr His1380Tyr Asp1682Tyr
BRCA2	BRCA1	Breast Cancer	1	aGAT-TAT TGT-TAT TGT-TAT aGAT-TAT aAAT-TAT TGT-TAT	Asp23Tyr Cys616Tyr Cys822Tyr Asn1420Tyr Asn1730Tyr Cys2605Tyr	TGT-TAT TGT-TAT TGC-TAC TGC-TAC TGT-TAT TGT-TAT tGAT-TAT tAAC-TAC gCAC-TAC TCT-TAT gCAT-TAT aGAT-TAT	Cys24Tyr Cys39Tyr Cys44Tyr Cys47Tyr Cys61Tyr Cys64Tyr Asp67Tyr Asn158Tyr His662Tyr Ser1253Tyr His1421Tyr Asp1739Tyr
BRCA2	BRCA1	Ovarian Cancer	1	aGAT-TAT aGAT-TAT	Asp23Tyr Asp1420Tyr	TGT-TAT TGC-TAC TGT-TAT tGAT-TAT aGAT-TAT	Cys24Tyr Cys47Tyr Cys61Tyr Asp67Tyr Asp1778Tyr
BRCA2	BRCA1	Prostate Cancer	0	TGC-TAC	Cys1290Tyr	tCAC-TAC	His448Tyr
CHEK2	BRCA2	Breast Cancer	1	gAAT-TAT gCAC-TAC	Asn352Tyr His371Tyr	aGAT-TAT TGT-TAT TGT-TAT aGAT-TAT aAAT-TAT TGT-TAT	Asp23Tyr Cys616Tyr Cys822Tyr Asp1420Tyr Asn1730Tyr Cys2605Tyr
CHEK2	ATM	Breast Cancer	0	gAAT-TAT gCAC-TAC	Asn352Tyr His371Tyr	TGT-TAT tCAT-TAT tCAT-TAT aGAT-TAT	Cys107Tyr His231Tyr His1380Tyr Asp1682Tyr
CHEK2	BRCA1	Breast Cancer	1	gAAT-TAT gCAC-TAC	Asn352Tyr His371Tyr	TGT-TAT TGT-TAT TGC-TAC TGC-TAC TGT-TAT TGT-TAT tGAT-TAT tAAC-TAC gCAC-TAC TCT-TAT gCAT-TAT aGAT-TAT	Cys24Tyr Cys39Tyr Cys44Tyr Cys47Tyr Cys61Tyr Cys64Tyr Asp67Tyr Asn158Tyr His662Tyr Ser1253Tyr His1421Tyr Asp1739Tyr
MSH2	MLH1	Bowel Cancer	1	tGAT-TAT TCT-TAT TGT-TAT gGAC-TAC aGAT-TAT gCAT-TAT cAAT-TAT aGAT-TAT	Asp283Tyr Ser323Tyr Cys333Tyr Asp506Tyr Asp603Tyr His639Tyr Asn671Tyr Asp748Tyr	TGT-TAT TGT-TAT cCAT-TAT aCAC-TAC	Cys39Tyr Cys77Tyr His264Tyr His718Tyr
MSH6	MLH1	Bowel Cancer	0	TACg-TAT	Tyr214Tyr	TGT-TAT TGT-TAT cCAT-TAT aCAC-TAC	Cys39Tyr Cys77Tyr His264Tyr His718Tyr
MSH6	MSH2	Bowel Cancer	0	TACg-TAT	Tyr214Tyr	tGAT-TAT TCT-TAT TGT-TAT gGAC-TAC aGAT-TAT gCAT-TAT cAAT-TAT aGAT-TAT	Asp283Tyr Ser323Tyr Cys333Tyr Asp506Tyr Asp603Tyr His639Tyr Asn671Tyr Asp748Tyr
PMS2	MLH1	Bowel Cancer	1	TGT-TAT	Cys843Tyr	TGT-TAT TGT-TAT cCAT-TAT aCAC-TAC	Cys39Tyr Cys77Tyr His264Tyr His718Tyr
Serine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
AR	RNASEL	Prostate Cancer	0	gACC-TCC	Thr559Ser	gGGC-AGC	Gly59Ser
ATM	NBN	Breast Cancer	1	tCCT-TCT aCCT-TCT tCCT-TCT AAT-AGT AAC-AGC aCCT-TCT tGGT-AGT AAC-AGC ACT-AGT	Pro604Ser Pro872Ser Pro884Ser Asn975Ser Asn2343Ser Pro2614Ser Gly2765Ser Asn3003Ser Thr2396Ser	aCCT-TCT	Pro199Ser
AXIN2	APC	Bowel Cancer	0	gGCT-TCT	Ala695Ser	tCCA-TCA TGT-TCT aCCT-TCT aGGT-AGT	Pro870Ser Cys947Ser Pro1467Ser Gly1921Ser

						AAC-AGC	Asn2593Ser
BARD1	MSH6	Breast Cancer	0	gCCC-TCC AAT-AGT AGGa-AGC AAT-AGT TGC-TCC	Pro24Ser Asn295Ser Arg378Ser Asn470Ser Cys557Ser	gGCC-TCC TTC-TCC	Ala25Ser Phe340Ser
BRCA1	MSH3	Bowel Cancer	0	gTGT-AGT	Cys39Ser	tCCA-TCA	Pro681Ser
BRCA1	BARD1	Breast Cancer	1	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser	gCCC-TCC AAT-AGT AGGa-AGC AAT-AGT TGC-TCC	Pro24Ser Asn295Ser Arg378Ser Asn470Ser Cys557Ser
BRCA1	MLH1	Bowel Cancer	0	gTGT-AGT	Cys39Ser	aGCT-TCT AAC-AGC AAT-AGT tGGC-AGC aGGT-AGT ATT-AGT tCGT-AGT TTG-TCG AAT-AGT AAC-AGC aCCC-TCC gCCC-TCC TTC-TCC AGGt-AGC	Ala29Ser Asn38Ser Asn64Ser Gly98Ser Gly101Ser Ile216Ser Arg265Ser Leu272Ser Asn338Ser Asn635Ser Pro640Ser Pro648Ser Phe656Ser Arg755Ser
BRCA1	MSH2	Bowel Cancer	0	gTGT-AGT	Cys39Ser	gGGC-AGC AAT-AGT gCCA-TCA cCCT-TCT AGAt-AGT TAT-TCT AAT-AGT AAT-AGT TCA-TGA tGGC-AGC aGGT-AGT TTT-TCT TCC-TTC ATT-AGT	Gly40Ser Asn127Ser Pro259Ser Pro336Ser Arg359Ser Tyr563Ser Asn583Ser Asn596Ser Ser612Term Gly669Ser Gly674Ser Phe694Ser Ser723Phe Ile884Ser
BRCA1	MSH6	Breast Cancer	1	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser	gGCC-TCC	Ala25Ser
BRCA1	ATM	Breast Cancer	1	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser	tCCT-TCT aCCT-TCT tCCT-TCT AAT-AGT AAC-AGC aCCT-TCT tGGT-AGT AAC-AGC ACT-AGT	Pro604Ser Pro872Ser Pro884Ser Asn975Ser Asn2343Ser Pro2614Ser Gly2765Ser Asn3003Ser Thr2396Ser
BRCA1	MSH6	Bowel Cancer	0	gTGT-AGT	Cys39Ser	TTC-TCC	Phe340Ser
						gGGC-AGC	Gly1139Ser
BRCA1	PALB2	Breast Cancer	1	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser	tCCC-TCC TTA-TCA AGGa-AGC tCCT-TCT tCCA-TCA	Pro5Ser Leu337Ser Arg365Ser Pro864Ser Pro918Ser

				tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser		
BRCA2	BARD1	Breast Cancer	1	tCCT-TCT gCCC-TCC AAT-AGT AAT-AGT gGCC-TCC TTA-TCA AAT-AGT	Pro222Ser Pro375Ser Asn588Ser Asn2706Ser Ala2717Ser Leu2805Ser Asn2930Ser	gCCC-TCC AAT-AGT AGGa-AGC AAT-AGT TGC-TCC	Pro24Ser Asn295Ser Arg378Ser Asn470Ser Cys557Ser
BRCA2	ATM	Breast Cancer	1	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT tGCC-TCC TGG-TCG TTT-TCT aGGT-AGT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser Ala1669Ser Trp1718Ser Phe1734Ser Gly1748Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser	tCCT-TCT aCCT-TCT tCCT-TCT AAT-AGT AAC-AGC aCCT-TCT tGGT-AGT AAC-AGC ACT-AGT	Pro604Ser Pro872Ser Pro884Ser Asn975Ser Asn2343Ser Pro2614Ser Gly2765Ser Asn3003Ser Thr2396Ser
BRCA2	PALB2	Breast Cancer	1	tCCT-TCT gCCC-TCC AAT-AGT AAT-AGT gGCC-TCC TTA-TCA AAT-AGT	Pro222Ser Pro375Ser Asn588Ser Asn2706Ser Ala2717Ser Leu2805Ser Asn2930Ser	tCCC-TCC TTA-TCA AGGa-AGC tCCT-TCT tCCA-TCA	Pro5Ser Leu337Ser Arg365Ser Pro864Ser Pro918Ser
BRCA2	BRCA1	Breast Cancer	1	tCCT-TCT gCCC-TCC AAT-AGT AAT-AGT gGCC-TCC TTA-TCA AAT-AGT	Pro222Ser Pro375Ser Asn588Ser Asn2706Ser Ala2717Ser Leu2805Ser Asn2930Ser	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser
BRCA2	BRCA1	Ovarian Cancer	1	aGGC-AGC gGCC-TCC TTA-TCA AAT-AGT	Gly2508Ser Ala2717Ser Leu2805Ser Asn2930Ser	gTGT-AGT tGGC-AGC ACT-AGT AGAg-AGT tGCC-TCC TGG-TCG aGGT-AGT ATC-AGC cACC-TCC gTGT-AGT	Cys61Ser Gly275Ser Thr1051Ser Arg1076Ser Ala1669Ser Trp1718Ser Gly1748Ser Ile1766Ser Thr1773Ser Cys1787Ser
CDKN2A	CDK4	Melanoma	0	gGGT-AGT gCCC-TCC ATC-AGC cGGC-AGC AAC-AGC gGGC-AGC gCCC-TCC	Gly23Ser Pro38Ser Ile49Ser Gly67Ser Asn71Ser Gly111Ser Pro114Ser	AAT-AGT	Asn41Ser
CHEK2	BRCA2	Breast Cancer	1	TAC-TCC aCCT-TCT	Tyr390Ser Pro426Ser	tCCT-TCT gCCC-TCC AAT-AGT AAT-AGT gGCC-TCC TTA-TCA AAT-AGT	Pro222Ser Pro375Ser Asn588Ser Asn2706Ser Ala2717Ser Leu2805Ser Asn2930Ser
CHEK2	ATM	Breast Cancer	1	TAC-TCC aCCT-TCT	Tyr390Ser Pro426Ser	tCCT-TCT aCCT-TCT tCCT-TCT AAT-AGT AAC-AGC aCCT-TCT tGGT-AGT AAC-AGC ACT-AGT	Pro604Ser Pro872Ser Pro884Ser Asn975Ser Asn2343Ser Pro2614Ser Gly2765Ser Asn3003Ser Thr2396Ser
CHEK2	NBN	Breast Cancer	1	TAC-TCC aCCT-TCT	Tyr390Ser Pro426Ser	aCCT-TCT	Pro199Ser
CHEK2	BRCA1	Breast Cancer	1	TAC-TCC aCCT-TCT	Tyr390Ser Pro426Ser	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser

						tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser
MET	CDH1	Stomach Cancer	1	cCCA-TCA	Pro1009Ser	aGGT-AGT tCCA-TCA aACT-TCT	Gly274Ser Pro429Ser Thr599Ser
MLH3	MLH1	Bowel Cancer	1	AAT-AGT cGGT-AGT AAT-AGT	Asn499Ser Gly981Ser Asn1007Ser	aGCT-TCT AAC-AGC AAT-AGT tGGC-AGC aGGT-AGT ATT-AGT tCGT-AGT TTG-TCG AAT-AGT AAC-AGC aCCC-TCC gCCC-TCC TTC-TCC AGGt-AGC	Ala29Ser Asn38Ser Asn64Ser Gly98Ser Gly101Ser Ile216Ser Arg265Ser Leu272Ser Asn338Ser Asn635Ser Pro640Ser Pro648Ser Phe656Ser Arg755Ser
MSH2	MLH1	Bowel Cancer	1	gGGC-AGC AAT-AGT gCCA-TCA cCCT-TCT AGAt-AGT AAT-AGT AAT-AGT tGGC-AGC aGGT-AGT TTT-TCT ATT-AGT	Gly40Ser Asn127Ser Pro259Ser Pro336Ser Arg359Ser Asn583Ser Asn596Ser Gly669Ser Gly674Ser Phe694Ser Ile884Ser	aGCT-TCT AAC-AGC AAT-AGT tGGC-AGC aGGT-AGT ATT-AGT tCGT-AGT TTG-TCG AAT-AGT AAC-AGC aCCC-TCC gCCC-TCC TTC-TCC AGGt-AGC	Ala29Ser Asn38Ser Asn64Ser Gly98Ser Gly101Ser Ile216Ser Arg265Ser Leu272Ser Asn338Ser Asn635Ser Pro640Ser Pro648Ser Phe656Ser Arg755Ser
MSH3	MLH1	Bowel Cancer	0	tCCA-TCA	Pro681Ser	aGCT-TCT AAC-AGC AAT-AGT tGGC-AGC aGGT-AGT ATT-AGT tCGT-AGT TTG-TCG AAT-AGT AAC-AGC aCCC-TCC gCCC-TCC TTC-TCC AGGt-AGC	Ala29Ser Asn38Ser Asn64Ser Gly98Ser Gly101Ser Ile216Ser Arg265Ser Leu272Ser Asn338Ser Asn635Ser Pro640Ser Pro648Ser Phe656Ser Arg755Ser
MSH3	MSH2	Bowel Cancer	1	tCCA-TCA	Pro681Ser	gGGC-AGC AAT-AGT gCCA-TCA cCCT-TCT AGAt-AGT TAT-TCT AAT-AGT AAT-AGT tGGC-AGC aGGT-AGT TTT-TCT ATT-AGT	Gly40Ser Asn127Ser Pro259Ser Pro336Ser Arg359Ser Tyr563Ser Asn583Ser Asn596Ser Gly669Ser Gly674Ser Phe694Ser Ile884Ser
MSH6	MLH1	Bowel Cancer	1	TTC-TCC gGGC-AGC	Phe340Ser Gly1139Ser	aGCT-TCT AAC-AGC AAT-AGT cAGT-GGT tGGC-AGC aGGT-AGT ATT-AGT tCGT-AGT TTG-TCG AAT-AGT AAC-AGC aCCC-TCC gCCC-TCC TTC-TCC AGGt-AGC	Ala29Ser Asn38Ser Asn64Ser Ser93Gly Gly98Ser Gly101Ser Ile216Ser Arg265Ser Leu272Ser Asn338Ser Asn635Ser Pro640Ser Pro648Ser Phe656Ser Arg755Ser
MSH6	MSH2	Bowel Cancer	1	gGGC-AGC	Gly1139Ser	gGGC-AGC AAT-AGT gCCA-TCA cCCT-TCT AGAt-AGT TAT-TCT AAT-AGT AAT-AGT tGGC-AGC aGGT-AGT TTT-TCT ATT-AGT	Gly40Ser Asn127Ser Pro259Ser Pro336Ser Arg359Ser Tyr563Ser Asn583Ser Asn596Ser Gly669Ser Gly674Ser Phe694Ser Ile884Ser
MSH6	MSH2	Bowel Cancer	1	gGGC-AGC	Gly1139Ser	gGGC-AGC AAT-AGT gCCA-TCA cCCT-TCT AGAt-AGT TAT-TCT AAT-AGT AAT-AGT tGGC-AGC aGGT-AGT TTT-TCT ATT-AGT	Gly40Ser Asn127Ser Pro259Ser Pro336Ser Arg359Ser Tyr563Ser Asn583Ser Asn596Ser Gly669Ser Gly674Ser Phe694Ser Ile884Ser
MUTYH	MSH6	Bowel Cancer	0	TAT-TCT AAC-AGC	Tyr180Ser Asn238Ser	TTC-TCC gGGC-AGC	Phe340Ser Gly1139Ser
MYC	BRCA1	Breast_Cancer	1	AAC-AGC	Asn26Ser	TTA-TCA gTGT-AGT tTGC-AGC gTGT-AGT TTA-TCA tGGC-AGC AGGc-AGC AAC-AGC	Leu22Ser Cys39Ser Cys44Ser Cys61Ser Leu63Ser Gly275Ser Arg331Ser Asn656Ser

						gCCA-TCA TTA-TCA ACT-AGT AGAg-AGT aCCT-TCT gGGT-AGT tGCC-TCC TGG-TCG TTT-TCT ATC-AGC cACC-TCC gTGT-AGT tGCT-TCT	Pro684Ser Leu892Ser Thr1051Ser Arg1076Ser Pro1150Ser Gly1201Ser Ala1669Ser Trp1718Ser Phe1734Ser Ile1766Ser Thr1773Ser Cys1787Ser Ala1789Ser
XRCC2	RAD51C	Breast Cancer	0	tGCC-TCC	Ala16Ser	TTA-TCA	Leu219Ser

Aspartate

Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
ATM	TP53	Breast Cancer	0	GGT-GAT aAAT-GAT aCAT-GAT GGT-GAT	Gly301Asp Asn1356Asp His2887Asp Gly2891Asp	GGC-GAC	Gly244Asp
BRCA1	TP53	Breast Cancer	1	GGC-GAC cTAT-GAT GAAg-GAC GGC-GAC GAGg-GAC GAAg-GAC GGT-GAT GGT-GAT GAGc-GAT	Gly275Asp Tyr465Asp Glu827Asp Gly960Asp Glu1219Asp Glu1581Asp Gly1748Asp Gly1788Asp Glu1794Asp	GGC-GAC	Gly244Asp
BRCA1	CREBBP	Ovarian Cancer	0	GGT-GAT GTC-GAC	Gly1788Asp Val1804Asp	cAAC-GAC	Asn1978Asp
BRCA1	ATM	Breast Cancer	1	GGC-GAC cTAT-GAT GAAg-GAC GGC-GAC GAGg-GAC GAAg-GAC GGT-GAT GGT-GAT GAGc-GAT	Gly275Asp Tyr465Asp Glu827Asp Gly960Asp Glu1219Asp Glu1581Asp Gly1748Asp Gly1788Asp Glu1794Asp	GGT-GAT aAAT-GAT aCAT-GAT GGT-GAT	Gly301Asp Asn1356Asp His2887Asp Gly2891Asp
BRCA1	PALB2	Breast Cancer	1	GGC-GAC cTAT-GAT GAAg-GAC GGC-GAC GAGg-GAC GAAg-GAC GGT-GAT GGT-GAT GAGc-GAT	Gly275Asp Tyr465Asp Glu827Asp Gly960Asp Glu1219Asp Glu1581Asp Gly1748Asp Gly1788Asp Glu1794Asp	GAGg-GAC GAGt-GAC	Glu1018Asp Glu1083Asp
BRCA2	ATM	Breast Cancer	1	GGC-GAC cTAT-GAT GAAg-GAC GGC-GAC GAGg-GAC GAAg-GAC GGT-GAT GGT-GAT GAGc-GAT	Gly275Asp Tyr465Asp Glu827Asp Gly960Asp Glu1219Asp Glu1581Asp Gly1748Asp Gly1788Asp Glu1794Asp	GGT-GAT aAAT-GAT aCAT-GAT GGT-GAT	Gly301Asp Asn1356Asp His2887Asp Gly2891Asp
BRCA2	TP53	Breast Cancer	1	GGT-GAT GGC-GAC aTAT-GAT GGT-GAT	Gly1771Asp Gly2544Asp Tyr2660Asp Gly2748Asp	GGC-GAC	Gly244Asp
BRCA2	BRCA1	Breast Cancer	1	GGT-GAT GGC-GAC aTAT-GAT GGT-GAT	Gly1771Asp Gly2544Asp Tyr2660Asp Gly2748Asp	GGC-GAC cTAT-GAT GAAg-GAC GGC-GAC GAGg-GAC GAAg-GAC GGT-GAT GGT-GAT GAGc-GAT	Gly275Asp Tyr465Asp Glu827Asp Gly960Asp Glu1219Asp Glu1581Asp Gly1748Asp Gly1788Asp Glu1794Asp
BRCA2	BRCA1	Ovarian Cancer	1	GGT-GAT GGC-GAC GGT-GAT GGT-GAT	Gly1771Asp Gly2544Asp Gly2748Asp Gly2901Asp	GGT-GAT GTC-GAC	Gly1788Asp Val1804Asp
BRCA2	PALB2	Breast Cancer	0	GGT-GAT GGC-GAC aTAT-GAT GGT-GAT	Gly1771Asp Gly2544Asp Tyr2660Asp Gly2748Asp	GAGg-GAC GAGt-GAC	Glu1018Asp Glu1083Asp
MSH6	MLH1	Bowel Cancer	1	GAAg-GAC GCT-GAT tCAC-GAC	Glu619Asp Ala1021Asp His1248Asp	GAAg-GAT gAAC-GAC GGC-GAC GGT-GAT GAGg-GAT GCT-GAT GTT-GAT GGT-GAT gTAT-GAT gTAC-GAC GTT-GAT GCC-GAC GCC-GAC GCC-GAC GCT-GAT GAGg-GAT	Glu23Asp Asn38Asp Gly65Asp Gly101Asp Glu102Asp Ala111Asp Val113Asp Gly244Asp Tyr280Asp Tyr293Asp Val384Asp Ala539Asp Ala586Asp Ala589Asp Ala608Asp Glu663Asp
MSH6	MSH2	Bowel Cancer	0	GAAg-GAC GCT-GAT tCAC-GAC	Glu619Asp Ala1021Asp His1248Asp	GGT-GAT GTT-GAT GTT-GAT gTAT-GAT GGC-GAC GAAa-GAT GGC-GAC GGT-GAT GAGc-GAC	Gly149Asp Val161Asp Val163Asp Tyr165Asp Gly322Asp Glu464Asp Gly669Asp Gly674Asp Glu878Asp
PMS2	MSH2	Bowel Cancer	1	GGC-GAC	Gly750Asp	GGT-GAT GTT-GAT GTT-GAT gTAT-GAT GGC-GAC GAAa-GAT	Gly149Asp Val161Asp Val163Asp Tyr165Asp Gly322Asp Glu464Asp

						GGC-GAC GGT-GAT GAGc-GAC	Gly669Asp Gly674Asp Glu878Asp
PMS2	MLH1	Bowel Cancer	1	GGC-GAC	Gly750Asp	GAAg-GAT gAAC-GAC GGC-GAC GGT-GAT GAGg-GAT GCT-GAT GTT-GAT GGT-GAT gTAT-GAT gTAC-GAC GTT-GAT GCC-GAC GCC-GAC GCC-GAC GCT-GAT GAGg-GAT	Glu23Asn Asn38Asn Gly65Asn Gly101Asn Glu102Asn Ala111Asn Val113Asn Gly244Asn Tyr280Asn Tyr293Asn Val384Asn Ala539Asn Ala586Asn Ala589Asn Ala608Asn Glu663Asn
TP53	MSH2	Bowel Cancer	1	GGC-GAC	Gly244Asp	GGT-GAT GTT-GAT GTT-GAT gTAT-GAT GGC-GAC GAAa-GAT GGC-GAC GGT-GAT GAGc-GAC	Gly149Asp Val161Asp Val163Asp Tyr165Asp Gly322Asp Glu464Asp Gly669Asp Gly674Asp Glu878Asp
Cysteine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	1	cCGT-TGT tCGT-TGT TTC-TGC aCGT-TGT tCGT-TGT TTT-TGT TCT-TGT tCGC-TGC gCGC-TGC	Arg13Cys Arg337Cys Phe627Cys Arg720Cys Arg981Cys Phe1463Cys Ser2592Cys Arg2691Cys Arg2854Cys	gCGC-TGC	Arg181Cys
ATM	TOPBP1	Breast Cancer	1	cCGT-TGT tCGT-TGT TTC-TGC aCGT-TGT tCGT-TGT TTT-TGT TCT-TGT tCGC-TGC gCGC-TGC	Arg13Cys Arg337Cys Phe627Cys Arg720Cys Arg981Cys Phe1463Cys Ser2592Cys Arg2691Cys Arg2854Cys	tCGT-TGT	Arg309Cys
BRCA1	TP53	Breast Cancer	1	TAT-TGT TAC-TGC TGCa-TGT gCGT-TGT gGGT-TGT tAGT-TGT gCGT-TGT TTC-TGC gCGC-TGC TCT-TGT TAC-TGC TAC-TGC tAGC-TGC TAC-TGC	Tyr105Cys Tyr179Cys Cys197Cys Arg252Cys Gly263Cys Ser264Cys Arg496Cys Phe861Cys Arg866Cys Ser1241Cys Tyr1522Cys Tyr1666Cys Ser1715Cys Tyr1853Cys	gCGC-TGC	Arg181Cys
BRCA1	TOPBP1	Breast Cancer	1	TAT-TGT TAC-TGC TGCa-TGT gCGT-TGT gGGT-TGT tAGT-TGT gCGT-TGT TTC-TGC gCGC-TGC TCT-TGT TAC-TGC TAC-TGC tAGC-TGC TAC-TGC	Tyr105Cys Tyr179Cys Cys197Cys Arg252Cys Gly263Cys Ser264Cys Arg496Cys Phe861Cys Arg866Cys Ser1241Cys Tyr1522Cys Tyr1666Cys Ser1715Cys Tyr1853Cys	tCGT-TGT	Arg309Cys
BRCA1	TOPBP1	Ovarian Cancer	1	TGCa-TGT gCGT-TGT TTC-TGC gCGC-TGC gAGT-TGT tAGC-TGC TAC-TGC	Cys197Cys Arg496Cys Phe861Cys Arg866Cys Ser1613Cys Ser1715Cys Tyr1853Cys	tCGT-TGT	Arg309Cys
BRCA1	PALB2	Breast Cancer	1	TAT-TGT TAC-TGC TGCa-TGT gCGT-TGT gGGT-TGT tAGT-TGT gCGT-TGT TTC-TGC gCGC-TGC TCT-TGT TAC-TGC TAC-TGC tAGC-TGC TAC-TGC	Tyr105Cys Tyr179Cys Cys197Cys Arg252Cys Gly263Cys Ser264Cys Arg496Cys Phe861Cys Arg866Cys Ser1241Cys Tyr1522Cys Tyr1666Cys Ser1715Cys Tyr1853Cys	TAC-TGC	Tyr28Cys
BRCA2	ATM	Breast Cancer	1	TAT-TGT TAC-TGC gCGT-TGT gGGT-TGT tAGT-TGT gCGT-TGT TTC-TGC gCGC-TGC TCT-TGT TAC-TGC gAGT-TGT TAC-TGC tAGC-TGC TAC-TGC	Tyr105Cys Tyr179Cys Arg252Cys Gly263Cys Ser264Cys Arg496Cys Phe861Cys Arg866Cys Ser1241Cys Tyr1522Cys Ser1613Cys Tyr1666Cys Ser1715Cys Tyr1853Cys	cCGT-TGT tCGT-TGT TTC-TGC aCGT-TGT tCGT-TGT TTT-TGT TCT-TGT tCGC-TGC gCGC-TGC	Arg13Cys Arg337Cys Phe627Cys Arg720Cys Arg981Cys Phe1463Cys Ser2592Cys Arg2691Cys Arg2854Cys

BRCA2	BRCA1	Breast Cancer	1	TAT-TGT TCT-TGT aCGT-TGT TCC-TGC TAT-TGT tCGT-TGT TAC-TGC TAT-TGT aCGC-TGC TAT-TGT TAT-TGT gAGT-TGT	Tyr42Cys Ser1261Cys Arg2034Cys Ser2072Cys Tyr2094Cys Arg2108Cys Tyr2222Cys Tyr2363Cys Arg2502Cys Tyr3035Cys Tyr3092Cys Ser3303Cys	TAT-TGT TAC-TGC gCGT-TGT gGGT-TGT tAGT-TGT gCGT-TGT TTC-TGC gCGC-TGC TCT-TGT TAC-TGC TAC-TGC tAGC-TGC TAC-TGC	Tyr105Cys Tyr179Cys Arg252Cys Gly263Cys Ser264Cys Arg496Cys Phe861Cys Arg866Cys Ser1241Cys Tyr1522Cys Tyr1666Cys Ser1715Cys Tyr1853Cys
BRCA2	BRCA1	Ovarian Cancer	1	aCGT-TGT tCGT-TGT aCGC-TGC	Arg2034Cys Arg2108Cys Arg2502Cys	TGCa-TGT gCGT-TGT TTC-TGC gCGC-TGC gAGT-TGT tAGC-TGC TAC-TGC	Cys197Cys Arg496Cys Phe861Cys Arg866Cys Ser1613Cys Ser1715Cys Tyr1853Cys
BRCA2	PALB2	Breast Cancer	1	TAT-TGT TCT-TGT aCGT-TGT TCC-TGC TAT-TGT tCGT-TGT TAC-TGC TAT-TGT aCGC-TGC TAT-TGT TAT-TGT gAGT-TGT	Tyr42Cys Ser1261Cys Arg2034Cys Ser2072Cys Tyr2094Cys Arg2108Cys Tyr2222Cys Tyr2363Cys Arg2502Cys Tyr3035Cys Tyr3092Cys Ser3303Cys	TAC-TGC	Tyr28Cys
BRCA2	TP53	Breast Cancer	1	TAT-TGT TCT-TGT aCGT-TGT TCC-TGC TAT-TGT tCGT-TGT TAC-TGC TAT-TGT aCGC-TGC TAT-TGT TAT-TGT gAGT-TGT	Tyr42Cys Ser1261Cys Arg2034Cys Ser2072Cys Tyr2094Cys Arg2108Cys Tyr2222Cys Tyr2363Cys Arg2502Cys Tyr3035Cys Tyr3092Cys Ser3303Cys	gCGC-TGC	Arg181Cys
CDKN2A	CDK4	Melanoma	1	gGGT-TGT gCGC-TGC tCGC-TGC	Gly23Cys Arg107Cys Arg124Cys	cCGT-TGT	Arg24Cys
CHEK2	BRCA2	Breast Cancer	1	cCGT-TGT	Arg346Cys	TAT-TGT TCT-TGT aCGT-TGT TCC-TGC TAT-TGT tCGT-TGT TAC-TGC TAT-TGT aCGC-TGC TAT-TGT TAT-TGT gAGT-TGT	Tyr42Cys Ser1261Cys Arg2034Cys Ser2072Cys Tyr2094Cys Arg2108Cys Tyr2222Cys Tyr2363Cys Arg2502Cys Tyr3035Cys Tyr3092Cys Ser3303Cys
CHEK2	BRCA2	Breast Cancer	1	cCGT-TGT	Arg346Cys	TAT-TGT TCT-TGT aCGT-TGT TCC-TGC TAT-TGT tCGT-TGT TAC-TGC TAT-TGT aCGC-TGC TAT-TGT TAT-TGT gAGT-TGT	Tyr42Cys Ser1261Cys Arg2034Cys Ser2072Cys Tyr2094Cys Arg2108Cys Tyr2222Cys Tyr2363Cys Arg2502Cys Tyr3035Cys Tyr3092Cys Ser3303Cys
CHEK2	ATM	Breast Cancer	1	cCGT-TGT	Arg346Cys	cCGT-TGT tCGT-TGT TTC-TGC aCGT-TGT tCGT-TGT TTT-TGT TCT-TGT tCGC-TGC gCGC-TGC	Arg13Cys Arg337Cys Phe627Cys Arg720Cys Arg981Cys Phe1463Cys Ser2592Cys Arg2691Cys Arg2854Cys
CHEK2	TP53	Breast Cancer	0	cCGT-TGT	Arg346Cys	gCGC-TGC	Arg181Cys
CHEK2	BRCA1	Breast Cancer	1	cCGT-TGT	Arg346Cys	TAT-TGT TAC-TGC TGCa-TGT gCGT-TGT gGGT-TGT tAGT-TGT gCGT-TGT TTC-TGC gCGC-TGC TCT-TGT TAC-TGC TAC-TGC tAGC-TGC TAC-TGC	Tyr105Cys Tyr179Cys Cys197Cys Arg252Cys Gly263Cys Ser264Cys Arg496Cys Phe861Cys Arg866Cys Ser1241Cys Tyr1522Cys Tyr1666Cys Ser1715Cys Tyr1853Cys
MLH3	MLH1	Bowel Cancer	1	aCGT-TGT tGGT-TGT	Arg647Cys Gly933Cys	cCGC-TGC tCGC-TGC tCGT-TGT TAT-TGT tCGT-TGT TCT-TGT TAT-TGT gCGC-TGC	Arg18Cys Arg217Cys Arg265Cys Tyr379Cys Arg385Cys Ser420Cys Tyr646Cys Arg725Cys
MSH2	MLH1	Bowel Cancer	1	TAT-TGT TAT-TGT TCT-TGT tGGT-TGT TAT-TGT	Tyr43Cys Tyr98Cys Ser323Cys Gly548Cys Tyr619Cys	cCGC-TGC tCGC-TGC tCGT-TGT TAT-TGT tCGT-TGT TCT-TGT TAT-TGT gCGC-TGC	Arg18Cys Arg217Cys Arg265Cys Tyr379Cys Arg385Cys Ser420Cys Tyr646Cys Arg725Cys
MSH6	MLH1	Bowel Cancer	1	TCC-TGC TAC-TGC	Ser503Cys Tyr850Cys	cCGC-TGC tCGC-TGC	Arg18Cys Arg217Cys

				TAT-TGT tCGC-TGC	Tyr969Cys Arg1076Cys	tCGT-TGT TAT-TGT tCGT-TGT TCT-TGT TAT-TGT gCGC-TGC	Arg265Cys Tyr379Cys Arg385Cys Ser420Cys Tyr646Cys Arg725Cys
MSH6	MSH2	Bowel Cancer	1	TAC-TGC TAT-TGT tCGC-TGC	Tyr850Cys Tyr969Cys Arg1076Cys	TAT-TGT TAT-TGT TCT-TGT tGGT-TGT TAT-TGT	Tyr43Cys Tyr98Cys Ser323Cys Gly548Cys Tyr619Cys
MUTYH	MSH6	Bowel Cancer	0	cCGT-TGT	Arg245Cys	TAC-TGC TAT-TGT tCGC-TGC	Tyr850Cys Tyr969Cys Arg1076Cys
OGG1	XPC	Lung Cancer	0	TCC-TGC	Ser326Cys	tCGT-TGT	Arg671Cys
TP53	MSH2	Gastric Cancer	0	gCGC-TGC	Arg283Cys	TAT-TGT	Tyr408Cys
XRCC2	RAD51C	Breast Cancer	0	TGGc-TGT	Trp231Cys	tCGT-TGT	Arg249Cys
				Glutamine			
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
AR	SMAD4	Breast Cancer	0	CGA-CAA	Arg608Gln	CAGg-CAA	Gln450Gln
AR	BRCA1	Breast Cancer	1	CGA-CAA	Arg608Gln	cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Lys519Gln Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln
AR	TP53	Breast Cancer	1	CGA-CAA	Arg608Gln	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln
ATM	TP53	Breast Cancer	1	CGA-CAA CGG-CAG CGA-CAA	Arg805Gln Arg924Gln Arg250Gln	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln
ATM	RAD51	Breast Cancer	1	CGA-CAA CGG-CAG CGA-CAA	Arg805Gln Arg924Gln Arg250Gln	CGG-CAG	Arg150Gln
BRCA1	RAD51	Breast Cancer	1	cAAG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Lys519Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln	CGG-CAG	Arg150Gln
BRCA1	AURKA	Breast Cancer	0	GATg-GAG aAAG-GAG cAAA-GAA gAAA-GAA GTG-GAG GGA-GAA GCG-GAG GGA-GAA GATg-GAG GTG-GAG	Asp67Glu Lys654Glu Lys820Glu Lys1606Glu Val1696Glu Gly1706Glu Ala1708Glu Gly1738Glu Asp1739Glu Val1838Glu	GTA-GAA	Val377Glu
BRCA1	TP53	Breast Cancer	1	CAGg-CAA cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Gln262Gln Lys519Gln Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln
BRCA1	SMAD4	Breast Cancer	1	CAGg-CAA cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Gln262Gln Lys519Gln Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln	CAGg-CAA	Gln450Gln
BRCA1	BRIP1	Breast Cancer	0	GATg-GAG aAAG-GAG cAAA-GAA gAAA-GAA GTG-GAG GGA-GAA GCG-GAG GGA-GAA GATg-GAG GTG-GAG	Asp67Glu Lys654Glu Lys820Glu Lys1606Glu Val1696Glu Gly1706Glu Ala1708Glu Gly1738Glu Asp1739Glu Val1838Glu	cCAG-GAG	Gln944Glu
BRCA1	PALB2	Breast Cancer	0	CAGg-CAA cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Gln262Gln Lys519Gln Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln	aGAA-CAA	Glu672Gln
BRCA1	ATM	Breast Cancer	1	CAGg-CAA cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Gln262Gln Lys519Gln Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln	CGA-CAA CGG-CAG CGA-CAA	Arg805Gln Arg924Gln Arg250Gln
BRCA2	ATM	Breast Cancer	1	CAGg-CAA cAAG-CAG CGG-CAG cAAG-CAG CGA-CAA CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Gln262Gln Lys519Gln Arg841Gln Lys918Gln Arg1203Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln	CGA-CAA CGG-CAG CGA-CAA	Arg805Gln Arg924Gln Arg250Gln
BRCA2	RAD51	Breast Cancer	1	cGAG-CAG tAAA-CAA CGA-CAA CGG-CAG CGG-CAG	Glu1895Gln Lys2150Gln Arg2318Gln Arg2784Gln Arg3052Gln	CGG-CAG	Arg150Gln
BRCA2	BRCA1	Breast Cancer	1	cGAG-CAG tAAA-CAA	Glu1895Gln Lys2150Gln	CAGg-CAA cAAG-CAG	Gln262Gln Lys519Gln

				CGA-CAA CGG-CAG CGG-CAG	Arg2318Gln Arg2784Gln Arg3052Gln	CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln
BRCA2	BRCA1	Ovarian Cancer	1	CGA-CAA	Arg2318Gln	cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA CGG-CAG CGA-CAA	Lys519Gln Arg841Gln Lys918Gln His1686Gln Arg1699Gln Arg1751Gln
BRCA2	BRCA1	Prostate Cancer	0	CCA-CAA	Pro2347Gln	CGA-CAA	Arg1203Gln
BRCA2	PALB2	Breast Cancer	0	cGAG-CAG tAAA-CAA CGA-CAA CGG-CAG CGG-CAG	Glu1895Gln Lys2150Gln Arg2318Gln Arg2784Gln Arg3052Gln	aGAA-CAA	Glu672Gln
BRCA2	TP53	Breast Cancer	1	cGAG-CAG tAAA-CAA CGA-CAA CGG-CAG CGG-CAG	Glu1895Gln Lys2150Gln Arg2318Gln Arg2784Gln Arg3052Gln	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln
CHD1L	XRCC1	Lung Cancer	0	CACc-CAG	His350Gln	CGG-CAG	Arg399Gln
CHEK2	MSH2	Bowel Cancer	1	CGG-CAG	Arg145Gln	CCG-CAG CACg-CAG CGG-CAG gAAA-CAA CATg-CAA	Pro5Gln His46Gln Arg243Gln Lys246Gln His639Gln
CHEK2	TP53	Breast Cancer	1	CGA-CAA	Arg137Gln	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln
CHEK2	BRCA1	Breast Cancer	1	CGA-CAA	Arg137Gln	cAAG-CAG CGG-CAG cAAG-CAG CATg-CAA gAAA-CAA CGG-CAG CGA-CAA	Lys519Gln Arg841Gln Lys918Gln His1686Gln Lys1690Gln Arg1699Gln Arg1751Gln
EXO1	MSH2	Bowel Cancer	2	GGA-GAA	Gly759Glu	GTA-GAA cCAA-GAA GGA-GAA tAAA-GAA GTA-GAA	Val470Glu Gln690Glu Gly759Glu Lys845Glu Val923Glu
MLH3	MLH1	Bowel Cancer	0	aGAA-CAA	Glu624Gln	aGAG-CAG CGA-CAA cAAA-CAA CGG-CAG cAAA-CAA CGG-CAG CGA-CAA CTA-CAA	Glu199Gln Arg226Gln Lys286Gln Arg389Gln Lys443Gln Arg474Gln Arg659Gln Leu749Gln
MSH2	MLH1	Gastric Cancer	0	tCAA-GAA	Gln824Glu	CATg-CAG tGAA-CAA	His109Gln Glu433Gln
MSH2	MLH1	Bowel Cancer	1	CCG-CAG CACg-CAG CGG-CAG gAAA-CAA CATg-CAA	Pro5Gln His46Gln Arg243Gln Lys246Gln His639Gln	aGAG-CAG CGA-CAA cAAA-CAA CGG-CAG cAAA-CAA CGG-CAG CGA-CAA CTA-CAA	Glu199Gln Arg226Gln Lys286Gln Arg389Gln Lys443Gln Arg474Gln Arg659Gln Leu749Gln
MSH3	MLH1	Bowel Cancer	1	CGG-CAG	Arg949Gln	aGAG-CAG CGA-CAA cAAA-CAA CGG-CAG cAAA-CAA CGG-CAG CGA-CAA CTA-CAA	Glu199Gln Arg226Gln Lys286Gln Arg389Gln Lys443Gln Arg474Gln Arg659Gln Leu749Gln
MSH3	MSH2	Bowel Cancer	1	CGG-CAG	Arg949Gln	CCG-CAG CACg-CAG CGG-CAG gAAA-CAA CATg-CAA	Pro5Gln His46Gln Arg243Gln Lys246Gln His639Gln
MSH6	MLH1	Bowel Cancer	0	tCAG-GAG CTG-CAG	Gln698Glu Leu1354Gln	CATg-CAG aGAG-CAG CGA-CAA cAAA-CAA CGG-CAG tGAA-CAA cAAA-CAA CGG-CAG CGA-CAA CTA-CAA	His109Gln Glu199Gln Arg226Gln Lys286Gln Arg389Gln Glu433Gln Lys443Gln Arg474Gln Arg659Gln Leu749Gln
MSH6	MSH2	Bowel Cancer	0	CTG-CAG	Leu1354Gln	CCG-CAG CACg-CAG CGG-CAG gAAA-CAA CATg-CAA	Pro5Gln His46Gln Arg243Gln Lys246Gln His639Gln
MUTYH	MSH6	Bowel Cancer	0	CGG-CAG	Arg184Gln	CTG-CAG	Leu1354Gln
PMS2	MSH2	Bowel Cancer	1	CGG-CAG	Arg20Gln	CCG-CAG CACg-CAG CGG-CAG gAAA-CAA CATg-CAA	Pro5Gln His46Gln Arg243Gln Lys246Gln His639Gln
PMS2	MLH1	Bowel Cancer	1	CGG-CAG	Arg20Gln	aGAG-CAG CGA-CAA cAAA-CAA CGG-CAG cAAA-CAA CGG-CAG CGA-CAA CTA-CAA	Glu199Gln Arg226Gln Lys286Gln Arg389Gln Lys443Gln Arg474Gln Arg659Gln Leu749Gln
PMS2	MSH3	Bowel Cancer	1	CGG-CAG	Arg20Gln	CGG-CAG	Arg949Gln
TP53	RAD51	Breast Cancer	1	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln	CGG-CAG	Arg150Gln
TP53	CDKN1A	Breast Cancer	1	CGA-CAA CGG-CAG	Arg306Gln Arg267Gln	CGA-CAA	Arg84Gln
Leucine							

Node1	Node2	Cancer	Match	Node1 codon change	Node1 Amino acid changes	Node2 codon change	Node2 Amino acid changes
AR	RNASEL	Prostate Cancer	0	CGC-CTC	Arg727Leu	tATC-CTC	Ile97Leu
ATM	TP53	Breast Cancer	1	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu	CCG-CTG CCT-CTT	Pro82Leu Pro278Leu
ATM	NBN	Breast Cancer	0	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu	CCG-CTG	Pro266Leu
BRCA1	FANCD2	Breast Cancer	0	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu	CTTt-CTG	Leu1366Leu
BRCA1	TP53	Breast Cancer	1	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu	CCG-CTG CCT-CTT	Pro82Leu Pro278Leu
BRCA1	RAD50	Breast Cancer	0	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu	tATA-CTA	Ile94Leu
BRCA1	RAD50	Ovarian Cancer	0	aGTA-CTA CCT-CTT CAT-CTT cATA-TTA CCA-CTA tGTG-CTG CGG-CTG	Val1116Leu Pro1238Leu His1421Leu Ile1593Leu Pro1637Leu Val1696Leu Arg1699Leu	tATA-CTA	Ile94Leu
BRCA1	ATM	Breast Cancer	0	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu
BRCA1	ATM	Breast Cancer	1	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu
BRCA1	PALB2	Breast Cancer	0	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu	CCC-CTC	Pro8Leu
BRCA2	FANCD2	Breast Cancer	0	gTTT-CTT aTTC-CTC CAG-CTG gCTC-GTC TTA-TGA TCG-TTG TCA-TTA aATG-TTG gGTA-TTA tTTT-CTT TCG-TTG	Phe32Leu Phe81Leu Gln713Leu Leu1019Val Leu1152Term Ser1172Leu Ser1271Leu Met1272Leu Val2166Leu Phe2234Leu Ser2670Leu	CTTt-CTG	Leu1366Leu
BRCA2	ATM	Breast Cancer	1	CCA-CTA aTTT-CTT TTTg-TTA CCG-CTG CCT-CTT aGTA-CTA CCT-CTT CAT-CTT cATA-TTA CCA-CTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro871Leu Pro1099Leu Val1116Leu Pro1238Leu His1421Leu Ile1593Leu Pro1637Leu Val1696Leu Arg1699Leu	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT TCA-TTA	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu Ser2394Leu
BRCA2	BRCA1	Breast Cancer	1	gTTT-CTT aTTC-CTC CAG-CTG TCG-TTG TCA-TTA aATG-TTG	Phe32Leu Phe81Leu Gln713Leu Ser1172Leu Ser1271Leu Met1272Leu	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu

				gGTA-TTA tTTT-CTT TCG-TTG	Val2166Leu Phe2234Leu Ser2670Leu	cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu
BRCA2	BRCA1	Ovarian Cancer	0	TCA-TTA tTTT-CTT TCG-TTG	Ser1271Leu Phe2234Leu Ser2670Leu	aGTA-CTA CCT-CTT CAT-CTT cATA-TTA CCA-CTA tGTG-CTG CGG-CTG	Val1116Leu Pro1238Leu His1421Leu Ile1593Leu Pro1637Leu Val1696Leu Arg1699Leu
BRCA2	PALB2	Breast Cancer	0	gTTT-CTT aTTC-CTC CAG-CTG TCG-TTG TCA-TTA gGTA-TTA TCG-TTG	Phe32Leu Phe81Leu Gln713Leu Ser1172Leu Ser1271Leu Val2166Leu Ser2670Leu	CCC-CTC	Pro8Leu
BRCA2	TP53	Breast Cancer	0	gTTT-CTT aTTC-CTC CAG-CTG TCG-TTG TCA-TTA aATG-TTG gGTA-TTA tTTT-CTT TCG-TTG	Phe32Leu Phe81Leu Gln713Leu Ser1172Leu Ser1271Leu Met1272Leu Val2166Leu Phe2234Leu Ser2670Leu	CCG-CTG CCT-CTT	Pro82Leu Pro278Leu
CHEK2	BRCA2	Breast Cancer	0	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	gTTT-CTT aTTC-CTC CAG-CTG TCG-TTG TCA-TTA aATG-TTG gGTA-TTA tTTT-CTT TCG-TTG	Phe32Leu Phe81Leu Gln713Leu Ser1172Leu Ser1271Leu Met1272Leu Val2166Leu Phe2234Leu Ser2670Leu
CHEK2	BRCA2	Breast Cancer	0	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	gTTT-CTT aTTC-CTC CAG-CTG TCG-TTG TCA-TTA aATG-TTG gGTA-TTA tTTT-CTT TCG-TTG	Phe32Leu Phe81Leu Gln713Leu Ser1172Leu Ser1271Leu Met1272Leu Val2166Leu Phe2234Leu Ser2670Leu
CHEK2	RAD50	Breast Cancer	0	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	tATA-CTA	Ile94Leu
CHEK2	ATM	Breast Cancer	0	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu
CHEK2	NBN	Breast Cancer	1	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	CCG-CTG	Pro266Leu
CHEK2	TP53	Breast Cancer	1	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	CCG-CTG CCT-CTT	Pro82Leu Pro278Leu
CHEK2	BRCA1	Breast Cancer	0	TTCt-TTA CCG-CTG	Phe427Leu Pro484Leu	CCA-CTA aTTT-CTT TTTg-TTA CCT-CTT aGTA-CTA CCT-CTT cATA-TTA TTTg-TTA tGTG-CTG CGG-CTG	Pro359Leu Phe461Leu Phe486Leu Pro1099Leu Val1116Leu Pro1238Leu Ile1593Leu Phe1695Leu Val1696Leu Arg1699Leu
EXO1	MLH1	Bowel Cancer	1	CCT-CTT CCG-CTG	Pro757Leu Pro770Leu	CCA-CTA CGA-CTA cATA-TTA CCC-CTC CAG-CTG CCG-CTG CCC-CTC CCC-CTC CCT-CTT CGA-CTA	Pro28Leu Arg226Leu Ile246Leu Pro496Leu Gln542Leu Pro581Leu Pro640Leu Pro648Leu Pro654Leu Arg659Leu
EXO1	PMS2	Bowel Cancer	0	CCT-CTT CCG-CTG	Pro757Leu Pro770Leu	CGA-CTA TCG-TTG	Arg563Leu Ser815Leu
EXO1	MSH2	Bowel Cancer	1	CCT-CTT CCG-CTG	Pro757Leu Pro770Leu	cATG-CTG cATG-TTG CCT-CTT TCA-TTA CGT-CTT aTTG-CTG CCA-CTA CCC-CTC CCA-CTA	Met1Leu Met1Leu Pro349Leu Ser473Leu Arg524Leu Leu556Leu Pro622Leu Pro670Leu Pro696Leu
FANCD2	NBN	Breast Cancer	0	CTTt-CTG	Leu1366Leu	CCG-CTG	Pro266Leu
FANCD2	ATM	Breast Cancer	0	CTTt-CTG	Leu1366Leu	tGTT-CTT aTTC-CTC tGTG-TTG CCT-CTT TCG-TTG aTTT-CTT	Val182Leu Phe582Leu Val613Leu Pro1112Leu Ser1383Leu Phe858Leu
MET	CDH1	Gastric Cancer	1	CCG-CTG	Pro791Leu	CCG-CTG cATA-CTA	Pro373Leu Ile415Leu
MSH2	MLH1	Bowel Cancer	1	cATG-CTG cATG-TTG CCT-CTT TCA-TTA CGT-CTT aTTG-CTG CCA-CTA CCC-CTC CCA-CTA	Met1Leu Met1Leu Pro349Leu Ser473Leu Arg524Leu Leu556Leu Pro622Leu Pro670Leu Pro696Leu	CCA-CTA CGA-CTA cATA-TTA CCC-CTC CAG-CTG CCG-CTG CCC-CTC CCC-CTC CCT-CTT CGA-CTA	Pro28Leu Arg226Leu Ile246Leu Pro496Leu Gln542Leu Pro581Leu Pro640Leu Pro648Leu Pro654Leu Arg659Leu
MSH6	MLH1	Bowel Cancer	1	TCG-TTG CCC-CTC gGTT-CTT	Ser580Leu Pro656Leu Val800Leu	CCA-CTA CGA-CTA cATA-TTA CCC-CTC CAG-CTG	Pro28Leu Arg226Leu Ile246Leu Pro496Leu Gln542Leu

						CCG-CTG CCC-CTC CCC-CTC CCT-CTT CGA-CTA	Pro581Leu Pro640Leu Pro648Leu Pro654Leu Arg659Leu
MSH6	MSH2	Bowel Cancer	1	TCG-TTG CCC-CTC gGTT-CTT	Ser580Leu Pro656Leu Val800Leu	cATG-CTG cATG-TTG CCT-CTT TCA-TTA CGT-CTT aTTG-CTG CCA-CTA CCC-CTC CCA-CTA	Met1Leu Met1Leu Pro349Leu Ser473Leu Arg524Leu Leu556Leu Pro622Leu Pro670Leu Pro696Leu
MSH6	MSH2	Bowel Cancer	1	TCG-TTG CCC-CTC gGTT-CTT	Ser580Leu Pro656Leu Val800Leu	cATG-CTG cATG-TTG CCT-CTT TCA-TTA CGT-CTT aTTG-CTG CCA-CTA CCC-CTC CCA-CTA	Met1Leu Met1Leu Pro349Leu Ser473Leu Arg524Leu Leu556Leu Pro622Leu Pro670Leu Pro696Leu
MUTYH	MSH6	Bowel Cancer	0	CCA-CTA	Pro295Leu	TCG-TTG CCC-CTC gGTT-CTT	Ser580Leu Pro656Leu Val800Leu
PMS2	MSH2	Bowel Cancer	0	CGA-CTA TCG-TTG	Arg563Leu Ser815Leu	cATG-CTG cATG-TTG CCT-CTT TCA-TTA CGT-CTT aTTG-CTG CCA-CTA CCC-CTC CCA-CTA	Met1Leu Met1Leu Pro349Leu Ser473Leu Arg524Leu Leu556Leu Pro622Leu Pro670Leu Pro696Leu
PMS2	MLH1	Bowel Cancer	1	CGA-CTA TCG-TTG	Arg563Leu Ser815Leu	CCA-CTA CGA-CTA cATA-TTA CCC-CTC CAG-CTG CCG-CTG CCC-CTC CCC-CTC CCT-CTT CGA-CTA	Pro28Leu Arg226Leu Ile246Leu Pro496Leu Gln542Leu Pro581Leu Pro640Leu Pro648Leu Pro654Leu Arg659Leu
TP53	MSH2	Bowel Cancer	1	CCC-CTC	Pro250Leu	CCT-CTT TCA-TTA CGT-CTT aTTG-CTG CCA-CTA CCC-CTC CCA-CTA	Pro349Leu Ser473Leu Arg524Leu Leu556Leu Pro622Leu Pro670Leu Pro696Leu
Phenylalanine							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	1	TCT-TTT TTGc-TTT cATT-TTT TGT-TTT tCTT-TTT	Ser151Phe Leu1574Phe Ile1804Phe Cys2488Phe Leu1420Phe	cCTC-TTC tCTT-TTT	Leu130Phe Leu194Phe
ATM	NBN	Breast Cancer	0	TCT-TTT TTGc-TTT cATT-TTT TGT-TTT tCTT-TTT	Ser151Phe Leu1574Phe Ile1804Phe Cys2488Phe Leu1420Phe	cCTT-TTT	Leu150Phe
BRCA1	TP53	Breast Cancer	1	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe	cCTC-TTC tCTT-TTT	Leu130Phe Leu194Phe
BRCA1	ATM	Breast Cancer	0	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe	TCT-TTT TTGc-TTT cATT-TTT TGT-TTT tCTT-TTT	Ser151Phe Leu1574Phe Ile1804Phe Cys2488Phe Leu1420Phe
BRCA1	NBN	Breast Cancer	0	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe	cCTT-TTT	Leu150Phe
BRCA1	MSH6	Breast Cancer	1	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe	TCT-TTT	Ser682Phe
BRCA1	PALB2	Breast Cancer	0	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe	aCTT-TTT	Leu100Phe
BRCA2	BRCA1	Breast Cancer	1	TTAg-TTT TCC-TTC TTGg-TTT TCT-TTT gATC-TTC TCT-TTT gATC-TTC	Leu24Phe Ser384Phe Leu1522Phe Ser2522Phe Ile2627Phe Ser2704Phe Ile2944Phe	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe
BRCA2	ATM	Breast Cancer	1	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe	TCT-TTT TTGc-TTT cATT-TTT TGT-TTT tCTT-TTT	Ser151Phe Leu1574Phe Ile1804Phe Cys2488Phe Leu1420Phe
BRCA2	BRCA1	Ovarian Cancer	1	TTAg-TTT tTTT-CTT TCT-TTT	Leu24Phe Phe2234Leu Ser2522Phe	TGC-TTC TGC-TTC TCT-TTT	Cys44Phe Cys47Phe Ser186Phe

				gATC-TTC aCTT-TTT	Ile2627Phe Leu2686Phe	TCT-TTT gGTT-TTT	Ser1655Phe Val1809Phe
BRCA2	PALB2	Breast Cancer	0	TTAg-TTT TCC-TTC TTGg-TTT TCT-TTT gATC-TTC TCT-TTT gATC-TTC	Leu24Phe Ser384Phe Leu1522Phe Ser2522Phe Ile2627Phe Ser2704Phe Ile2944Phe	aCTT-TTT	Leu100Phe
BRCA2	TP53	Breast Cancer	0	TTAg-TTT TCC-TTC TTGg-TTT TCT-TTT gATC-TTC TCT-TTT gATC-TTC	Leu24Phe Ser384Phe Leu1522Phe Ser2522Phe Ile2627Phe Ser2704Phe Ile2944Phe	cCTC-TTC tCTT-TTT	Leu130Phe Leu194Phe
CHEK2	BRCA2	Ovarian Cancer	1	TCT-TTT	Ser357Phe	TTAg-TTT TCT-TTT gATC-TTC aCTT-TTT	Leu24Phe Ser2522Phe Ile2627Phe Leu2686Phe
CHEK2	BRCA2	Breast Cancer	1	gCTT-TTT TCT-TTT	Leu174Phe Ser428Phe	TTAg-TTT TCC-TTC TTGg-TTT TCT-TTT gATC-TTC TCT-TTT gATC-TTC	Leu24Phe Ser384Phe Leu1522Phe Ser2522Phe Ile2627Phe Ser2704Phe Ile2944Phe
CHEK2	BRCA2	Prostate Cancer	0	cATC-TTC	Ile251Phe	cCTT-TTT	Leu3055Phe
CHEK2	ATM	Breast Cancer	1	gCTT-TTT TCT-TTT	Leu174Phe Ser428Phe	TCT-TTT TTGc-TTT cATT-TTT TGT-TTT tCTT-TTT	Ser151Phe Leu1574Phe Ile1804Phe Cys2488Phe Leu1420Phe
CHEK2	NBN	Breast Cancer	1	gCTT-TTT TCT-TTT	Leu174Phe Ser428Phe	cCTT-TTT	Leu150Phe
CHEK2	TP53	Breast Cancer	1	gCTT-TTT TCT-TTT	Leu174Phe Ser428Phe	cCTC-TTC tCTT-TTT	Leu130Phe Leu194Phe
CHEK2	BRCA1	Breast Cancer	1	gCTT-TTT TCT-TTT	Leu174Phe Ser428Phe	TGC-TTC TGC-TTC tCTC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Leu52Phe Ser186Phe Ser1655Phe Val1809Phe
CHEK2	BRCA1	Ovarian Cancer	1	TCT-TTT	Ser357Phe	TGC-TTC TGC-TTC TCT-TTT TCT-TTT gGTT-TTT	Cys44Phe Cys47Phe Ser186Phe Ser1655Phe Val1809Phe
MSH2	MLH1	Bowel Cancer	1	tCTT-TTT aCTT-TTT TGT-TTT TCC-TTC	Leu93Phe Leu390Phe Cys697Phe Ser723Phe	cATC-TTC tATC-TTC TCC-TTC TTAa-TTC cATT-TTT gCTC-TTC	Ile19Phe Ile25Phe Ser44Phe Leu166Phe Ile565Phe Leu582Phe
MSH6	MLH1	Bowel Cancer	0	TAT-TTT	Tyr397Phe	cATC-TTC tATC-TTC TCC-TTC TTAa-TTC cATT-TTT gCTC-TTC	Ile19Phe Ile25Phe Ser44Phe Leu166Phe Ile565Phe Leu582Phe
MSH6	MSH2	Breast Cancer	0	TAT-TTT	Tyr397Phe	tCTT-TTT aCTT-TTT TGT-TTT TCC-TTC	Leu93Phe Leu390Phe Cys697Phe Ser723Phe
Tryptophan							
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
BRCA1	TP53	Breast Cancer	1	TGTg-TGG TTG-TGG aCGG-TGG	Cys91Trp Leu523Trp Arg1699Trp	aCGG-TGG	Arg267Trp
BRCA1	MLH1	Breast Cancer	0	TGTg-TGG TTG-TGG	Cys91Trp Leu523Trp	cCGG-TGG	Arg389Trp
BRCA1	MLH1	Breast Cancer	1	aCGG-TGG	Arg1699Trp	cCGG-TGG	Arg389Trp
BRCA1	MSH2	Bowel Cancer	0	aCGG-TGG	Arg1699Trp	tGGG-TGG	Gly164Trp
CHEK2	BRCA2	Breast Cancer	1	tCGG-TGG	Arg3Trp	TGTg-TGG TGTc-TGG aCGG-TGG	Cys161Trp Cys554Trp Arg3052Trp
MSH2	MLH1	Bowel Cancer	1	tGGG-TGG	Gly164Trp	cGGG-TGG cCGG-TGG cCGG-TGG gAGG-TGG	Gly67Trp Arg389Trp Arg687Trp Arg755Trp
MSH3	MLH1	Bowel Cancer	0	TTG-TGG	Leu911Trp	cGGG-TGG cCGG-TGG cCGG-TGG gAGG-TGG	Gly67Trp Arg389Trp Arg687Trp Arg755Trp
MSH6	MLH1	Bowel Cancer	1	TGCc-TGG gCGG-TGG	Cys765Trp Arg772Trp	cGGG-TGG cCGG-TGG cCGG-TGG gAGG-TGG	Gly67Trp Arg389Trp Arg687Trp Arg755Trp
MSH6	MSH2	Bowel Cancer	0	gCGG-TGG	Arg772Trp	tGGG-TGG	Gly164Trp
BRCA1	MSH3	Bowel Cancer	0	aCGG-TGG	Arg1699Trp	TTG-TGG	Leu911Trp
BRCA1	MRE11A	Ovarian Cancer	1	TGTt-TGG tCGG-TGG	Cys644Trp Arg841Trp	gCGG-TGG	Arg305Trp
BRCA1	MSH6	Bowel Cancer	1	aCGG-TGG	Arg1699Trp	TGCc-TGG gCGG-TGG	Cys765Trp Arg772Trp
BRCA1	PALB2	Breast Cancer	1	TGTg-TGG TTG-TGG aCGG-TGG	Cys91Trp Leu523Trp Arg1699Trp	TTG-TGG	Leu939Trp
BRCA2	BRCA1	Breast Cancer	1	TGTg-TGG TGTc-TGG aCGG-TGG	Cys161Trp Cys554Trp Arg3052Trp	TGTg-TGG TTG-TGG aCGG-TGG	Cys91Trp Leu523Trp Arg1699Trp
BRCA2	BRCA1	Ovarian Cancer	1	TGTg-TGG	Cys161Trp	TGTt-TGG tCGG-TGG	Cys644Trp Arg841Trp
BRCA2	PALB2	Breast Cancer	0	TGTg-TGG TGTc-TGG aCGG-TGG	Cys161Trp Cys554Trp Arg3052Trp	TTG-TGG	Leu939Trp
BRCA2	TP53	Breast Cancer	1	TGTg-TGG	Cys161Trp	aCGG-TGG	Arg267Trp

				TGTc-TGG aCGG-TGG	Cys554Trp Arg3052Trp		
CHEK2	BRCA1	Breast Cancer	1	tCGG-TGG	Arg3Trp	TGTg-TGG TTG-TGG aCGG-TGG	Cys91Trp Leu523Trp Arg1699Trp
CHEK2	TP53	Breast Cancer	1	tCGG-TGG	Arg3Trp	aCGG-TGG	Arg267Trp
MSH3	MSH2	Bowel Cancer	0	TTG-TGG	Leu911Trp	tGGG-TGG	Gly164Trp
BRCA1	MLH1	Bowel Cancer	1	aCGG-TGG	Arg1699Trp	cCGG-TGG	Arg389Trp
				Valine			
Node1	**Node2**	**Cancer**	**Match**	**Node1 codon change**	**Node1 Amino acid changes**	**Node2 codon change**	**Node2 Amino acid changes**
ATM	TP53	Breast Cancer	0	aCTG-GTG tATA-GTA tCTG-GTG tTTG-GTG cCTT-GTT cATT-GTT cCTA-GTA GCA-GTA aATG-GTG GAT-GTT	Leu546Val Ile550Val Leu702Val Leu1255Val Leu1493Val Ile2030Val Leu2330Val Ala2466Val Met3011Val Asp1853Val	GCC-GTC	Ala138Val
BRCA1	ACACA	Ovarian Cancer	0	aATG-GTG tTTG-GTG cATG-GTG GAT-GTT GCG-GTG GAT-GTT GGT-GTT	Met1Val Leu246Val Met1400Val Asp1692Val Ala1708Val Asp1739Val Gly1788Val	GCG-GTG	Ala2271Val
BRCA1	AURKA	Breast Cancer	0	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val	GCA-GTA	Ala213Val
BRCA1	TP53	Breast Cancer	0	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val	GCC-GTC	Ala138Val
BRCA1	ATM	Breast Cancer	1	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val	aCTG-GTG tATA-GTA tCTG-GTG tTTG-GTG cCTT-GTT cATT-GTT cCTA-GTA GCA-GTA aATG-GTG GAT-GTT	Leu546Val Ile550Val Leu702Val Leu1255Val Leu1493Val Ile2030Val Leu2330Val Ala2466Val Met3011Val Asp1853Val
BRCA1	MSH6	Breast Cancer	0	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val	cATA-GTA	Ile251Val
BRCA2	ATM	Breast Cancer	1	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val	aCTG-GTG tATA-GTA tCTG-GTG tTTG-GTG cCTT-GTT cATT-GTT cCTA-GTA GCA-GTA aATG-GTG GAT-GTT	Leu546Val Ile550Val Leu702Val Leu1255Val Leu1493Val Ile2030Val Leu2330Val Ala2466Val Met3011Val Asp1853Val
BRCA2	FANCE	Breast Cancer	0	GAT-GTT GAT-GTT GTAa-GTT cATG-GTG gCTC-GTC aATG-GTG tCTT-GTT aATT-GTT GGC-GTC GCA-GTA GAA-GTA GAT-GTT cATA-GTA cATG-GTG aATT-GTT	Asp191Val Asp224Val Val465Val Met784Val Leu1019Val Met1272Val Leu1904Val Ile1929Val Gly2044Val Ala2466Val Glu2663Val Asp2723Val Ile2840Val Met2952Val Ile3412Val	tCTG-GTG	Leu1143Val
BRCA2	BRCA1	Breast Cancer	1	GAT-GTT GAT-GTT cATG-GTG gCTC-GTC aATG-GTG tCTT-GTT aATT-GTT GGC-GTC GCA-GTA GAA-GTA GAT-GTT cATA-GTA cATG-GTG aATT-GTT	Asp191Val Asp224Val Met784Val Leu1019Val Met1272Val Leu1904Val Ile1929Val Gly2044Val Ala2466Val Glu2663Val Asp2723Val Ile2840Val Met2952Val Ile3412Val	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val
BRCA2	BRCA1	Ovarian Cancer	1	GAT-GTT GTAa-GTT GAA-GTA GAT-GTT	Asp191Val Val465Val Glu2663Val Asp2723Val	aATG-GTG tTTG-GTG cATG-GTG GAT-GTT	Met1Val Leu246Val Met1400Val Asp1692Val

						GCG-GTG GAT-GTT GGT-GTT	Ala1708Val Asp1739Val Gly1788Val
BRCA2	TP53	Breast Cancer	0	GAT-GTT GAT-GTT GTAa-GTT cATG-GTG gCTC-GTC aATG-GTG tCTT-GTT aATT-GTT GGC-GTC GCA-GTA GAA-GTA GAT-GTT cATA-GTA cATG-GTG aATT-GTT	Asp191Val Asp224Val Val465Val Met784Val Leu1019Val Met1272Val Leu1904Val Ile1929Val Gly2044Val Ala2466Val Glu2663Val Asp2723Val Ile2840Val Met2952Val Ile3412Val	GCC-GTC	Ala138Val
CHEK2	BRCA2	Breast Cancer	1	aATT-GTT	Ile189Val	GAT-GTT cATG-GTG gCTC-GTC aATG-GTG tCTT-GTT aATT-GTT GGC-GTC GCA-GTA cATA-GTA cATG-GTG aATT-GTT	Asp224Val Met784Val Leu1019Val Met1272Val Leu1904Val Ile1929Val Gly2044Val Ala2466Val Ile2840Val Met2952Val Ile3412Val
CHEK2	ATM	Breast Cancer	1	aATT-GTT	Ile189Val	aCTG-GTG tATA-GTA tCTG-GTG tTTG-GTG cCTT-GTT GTT-GCT GTT-GCT cATT-GTT cCTA-GTA GTA-GCA GCA-GTA aATG-GTG GAT-GTT	Leu546Val Ile550Val Leu702Val Leu1255Val Leu1493Val Val1570Ala Val1729Ala Ile2030Val Leu2330Val Val2439Ala Ala2466Val Met3011Val Asp1853Val
CHEK2	TP53	Breast Cancer	0	aATT-GTT	Ile189Val	GCC-GTC	Ala138Val
CHEK2	BRCA1	Breast Cancer	1	aATT-GTT	Ile189Val	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val
FANCA	BRCA1	Breast Cancer	0	tCTG-GTG	Leu1143Val	aATG-GTG aATC-GTC tTTG-GTG GGT-GTT GGT-GTT GCG-GTG gATT-GTT cATG-GTG aATG-GTG GCG-GTG	Met1Val Ile21Val Leu246Val Gly263Val Gly552Val Ala622Val Ile1318Val Met1400Val Met1628Val Ala1708Val
MLH3	MLH1	Bowel Cancer	0	tATG-GTG	Met809Val	GCG-GTG gTTC-GTC GCT-GTT aCTG-GTG GCC-GTC cATC-GTC GGT-GTT cTTG-GTG GCA-GTA GAT-GTT GCT-GTT tCTG-GTG gCTT-GTT gCTC-GTC GCT-GTT tCTG-GTG GAT-GTT	Ala21Val Phe80Val Ala111Val Leu135Val Ala160Val Ile219Val Gly244Val Leu272Val Ala281Val Asp304Val Ala353Val Leu400Val Leu555Val Leu582Val Ala681Val Leu729Val Asp737Val
MSH2	MLH1	Bowel Cancer	2	cATG-GTG GCG-GTG GAC-GTC tCTT-GTT cATA-GTA gATA-GTA GCG-GTG aCTC-GTC GGT-GTT aCTT-GTT GAG-GTG GCT-GTT GGG-GTG GCT-GTT tATG-GTG	Met1Val Ala45Val Asp49Val Leu92Val Ile169Val Ile216Val Ala272Val Leu279Val Gly315Val Leu390Val Glu562Val Ala600Val Gly692Val Ala714Val Met813Val	GCG-GTG gTTC-GTC GCT-GTT aCTG-GTG GCC-GTC cATC-GTC GGT-GTT cTTG-GTG GCA-GTA GAT-GTT GCT-GTT tCTG-GTG gCTT-GTT gCTC-GTC GCT-GTT tCTG-GTG GAT-GTT	Ala21Val Phe80Val Ala111Val Leu135Val Ala160Val Ile219Val Gly244Val Leu272Val Ala281Val Asp304Val Ala353Val Leu400Val Leu555Val Leu582Val Ala681Val Leu729Val Asp737Val
MSH6	MLH1	Bowel Cancer	1	GCC-GTC cATA-GTA cATC-GTC gATG-GTG GCT-GTT GAA-GTA	Ala20Val Ile251Val Ile425Val Met492Val Ala787Val Glu1163Val	GCG-GTG gTTC-GTC GCT-GTT aCTG-GTG GCC-GTC cATC-GTC GGT-GTT cTTG-GTG GCA-GTA GAT-GTT GCT-GTT tCTG-GTG gCTT-GTT gCTC-GTC tATC-GTC	Ala21Val Phe80Val Ala111Val Leu135Val Ala160Val Ile219Val Gly244Val Leu272Val Ala281Val Asp304Val Ala353Val Leu400Val Leu555Val Leu582Val Ile655Val Ala681Val

						GCT-GTT tCTG-GTG GAT-GTT	Leu729Val Asp737Val
MSH6	MLH1	Womb Cancer	1	GCC-GTC cATA-GTA cATC-GTC gATG-GTG GCT-GTT GAA-GTA	Ala20Val Ile251Val Ile425Val Met492Val Ala787Val Glu1163Val	GCG-GTG gTTC-GTC GCT-GTT aCTG-GTG GCC-GTC cATC-GTC GGT-GTT cTTG-GTG GCA-GTA GAT-GTT GCT-GTT tCTG-GTG gCTT-GTT gCTC-GTC tATC-GTC GCT-GTT tCTG-GTG GAT-GTT	Ala21Val Phe80Val Ala111Val Leu135Val Ala160Val Ile219Val Gly244Val Leu272Val Ala281Val Asp304Val Ala353Val Leu400Val Leu555Val Leu582Val Ile655Val Ala681Val Leu729Val Asp737Val
MSH6	MSH2	Bowel Cancer	1	cATC-GTC gATG-GTG GCT-GTT GAA-GTA	Ile425Val Met492Val Ala787Val Glu1163Val	cATG-GTG GCG-GTG GAC-GTC tCTT-GTT cATA-GTA gATA-GTA GCG-GTG aCTC-GTC GGT-GTT aCTT-GTT GAG-GTG GCT-GTT GGG-GTG GCT-GTT tATG-GTG	Met1Val Ala45Val Asp49Val Leu92Val Ile169Val Ile216Val Ala272Val Leu279Val Gly315Val Leu390Val Glu562Val Ala600Val Gly692Val Ala714Val Met813Val
MUTYH	MSH6	Bowel Cancer	1	GCC-GTC	Ala373Val	GCC-GTC cATC-GTC gATG-GTG GCT-GTT GAA-GTA	Ala20Val Ile425Val Met492Val Ala787Val Glu1163Val
PMS2	MSH2	Bowel Cancer	2	cATG-GTG	Met1Val	cATG-GTG GCG-GTG GAC-GTC tCTT-GTT cATA-GTA gATA-GTA GCG-GTG aCTC-GTC GGT-GTT aCTT-GTT GAG-GTG GCT-GTT GGG-GTG GCT-GTT tATG-GTG	Met1Val Ala45Val Asp49Val Leu92Val Ile169Val Ile216Val Ala272Val Leu279Val Gly315Val Leu390Val Glu562Val Ala600Val Gly692Val Ala714Val Met813Val
PMS2	MLH1	Bowel Cancer	0	cATG-GTG	Met1Val	GCG-GTG gTTC-GTC GCT-GTT aCTG-GTG GCC-GTC cATC-GTC GGT-GTT cTTG-GTG GCA-GTA GAT-GTT GCT-GTT tCTG-GTG gCTT-GTT gCTC-GTC GCT-GTT tCTG-GTG GAT-GTT	Ala21Val Phe80Val Ala111Val Leu135Val Ala160Val Ile219Val Gly244Val Leu272Val Ala281Val Asp304Val Ala353Val Leu400Val Leu555Val Leu582Val Ala681Val Leu729Val Asp737Val
RAD23B	XPC	Lung Cancer	0	GCT-GTT	Ala249Val	GCG-GTG	Ala499Val
TP53	AURKA	Breast Cancer	0	GCC-GTC	Ala138Val	GCA-GTA	Ala213Val
XRCC2	RAD51C	Breast Cancer	0	aATT-GTT tTTT-GTT	Ile95Val Phe270Val	GGT-GTT aTTA-GTA	Gly125Val Leu262Val